Molecular Dynamics simulations of aqueous glutamate and the gly-pro-glu (GPE) tripeptide

Antonia Collis

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
to the
University of Edinburgh
October 2011

Abstract

Biomolecular systems, in particular those involving proteins and their constituents, have been the focus of much research in the last century. The relationship between experiment, development of models and simulation has enabled vast improvements in our knowledge of subjects such as protein folding and the processes by which key biomolecules affect the human body. In particular, vital information can be obtained from understanding the building blocks of polypeptides and proteins involved in these processes.

This work focuses on simulating two such building blocks; glutamate, the salt of the proteinogenic amino acid, glutamic acid, and glycine-proline-glutamate, or GPE, a related tripeptide. Both are important in neurotransmission processes in the brain. Glutamate is the most abundant neurotransmitter in the central nervous system and GPE is an important neuroprotective agent. This work aims to elucidate the key structural properties of aqueous solutions of glutamate and GPE, focusing on the solute-solute as well as the solute-solvent interactions. Both systems were considered with classical empirical potentials using the CHARMM22 force-field. The glutamate system has also been studied using Car-Parrinello Molecular Dynamics and classical parallel tempering.

In both the aqueous glutamate and GPE systems the molecules formed a large proportion of bifurcated bonding motifs with both carboxyl groups, but not with the amin (N-terminal) of the molecules. Bifurcated bonds form between solute molecules as well as in the solute-solvent interactions. The structure of the glutamate solution was found to be dependent on the initial configuration and thus the parallel tempering simulations enabled better sampling of the conformational landscape. In addition, in the glutamate system single water molecules form a stable structure by bonding to both the amine (N-terminal) and C_α-carboxyl within the same glutamate molecule.

Declaration

Except where otherwise stated, I declare that this thesis and the work described has been undertaken by myself unaided. The work reported has not been submitted in application for any previous degree or qualification.

A. Collis
18^{th} October 2011

Acknowledgements

I would like to express my sincere gratitude to Simon Bates and Paul Tulip for their enthusiasm, guidance and encouragement throughout my PhD. Without their supervision this work would not have been possible. I would also like to thank my parents, Glyn and Dawn Collis for all their encouragement and support over the last five years and Alana Collis for her help with Chemistry. Thanks also go to Grace Kim and Sophie Marion de Procé, for keeping me motivated, providing cups of coffee and a shoulder to cry on when everything went wrong. Finally, I would like to express my thanks to Daniel Holmes for, amongst a long list of things, keeping me sane throughout this experience, providing computer expertise and proof reading. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Proteins form one of the essential categories of biomolecules. Together with nucleic acids and lipids, they constitute the building-blocks of life and the functionality of the processes within all living organisms. Proteins provide the structural scaffold for cells and essential functions such as fighting disease through immune response, maintaining cellular lifecycles and catalytic reactions including metabolism. Understanding proteins, the amino acids that form them and the structure and dynamics of biological processes associated with them provides key insight into our understanding of life. The added incentive, if one were needed, is that by understanding biological structures we are better able to provide solutions to the failures of our bodies, by providing the information needed to create medicines to slow or even remove molecular degeneration, provide avenues of attacking tumour cells, unwanted DNA mutations, or simply providing the information required for better drug delivery by understanding how molecules interact with amino acids, peptides and DNA. The twenty amino acids involved in genetic code are characterised by properties such as hydrophobicity, hydrophilicity, size and the functional groups they have. The structure of proteins and protein-protein interactions are governed by these properties and the characteristic behaviour that the amino acids exhibit due to these physical attributes directly effect their functionality.

The study of biomolecules includes scientists from all disciplines; biology describes the cellular structure, chemistry considers the atomistic processes, physics provides a perspective from fundamental particles and their forces, and mathematics enables the formulation of models to explain observed phenomena.

In reality, the lines between these disciplines has become blurred in recent years with multi-disciplinary groups studying systems from a multitude of different perspectives.

Experimental observation facilitates the development of new models, which in turn provide insight that drives new experiments in a perpetual cycle of the desire to understand. Molecular modelling is fundamental to this process as it relies on the theoretical models. Molecular modelling can provide a test for theory as well as providing predictions unobtainable from experiment. Modelling enables scientists to consider a range of problems from understanding the equilibrium structure of simple small biologically representative molecules, through examining the processes of hydrogen bonding between groups of biomolecules, to investigating protein folding. We are currently in an era of exceptional growth in molecular modelling. The first molecular mechanics study was reported in 1946 as a theoretical computation by Hill, Westheimer and Meyer [8, 9]. Contemporary supercomputers are able to simulate tens of thousands of atoms and complete nano-second simulations in less than a week.

In this work computer simulations are used to gain an understanding of both the physical and chemical properties of biomolecular systems. Empirical force-fields, treating the systems under consideration atomistically with classical potentials, are used along with electronic structure calculations to understand the chemical aspects of the system from a physical force-driven perspective. The classical empirical forcefield approach enables the study of large systems for longer time-scales than electronic structure calculations. However, the results are limited by the dependence on the accuracy of the force-field. First principles simulations, using the electronic structure to facilitate an accurate computation of the inter and intra-molecular interactions, as the results are not dependant on a well-constructed force-field, but this approach is limited by the cost of the calculations restricting the simulations to small system sizes and short time-scales. This work considers both classical and quantum calculations, with the aim of overcoming the weaknesses of each method with the strengths of the other.

1.1 Biological systems and water

Water is as essential to life as proteins. The structure of proteins, which provide the scaffold for life, is often stabilised by the presence of water [10, 11, 12]. Water-peptide interactions on the molecular scale have been vitally important in understanding the role of water in the protein dynamics and bio-protein structures [13]. In particular, study of aqueous solutions of peptides and hydrophobic effects has enabled further understanding of protein folding mechanisms, self-aggregation [14] and the formation of secondary and tertiary structures [15].

Much of the research in the field of biomolecular simulation focuses on macromolecules and large-scale problems such as protein folding, or the structure of large protein fragments [16]. However, to best understand the mechanisms behind such events and the structural integrity of the large systems it is prudent to consider the constituents also. As such, this work takes a 'bottom-up' approach. Previous work has considered hydrogen-bond interactions in pure water and around hydrophobic and hydrophilic groups [17, 18, 19, 20, 21, 22]. More recently amino acids themselves and dipeptides have been considered both experimentally and through computational experiment (for a non-exhaustive list of examples see refs [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33]). These experiments and simulations have included examining the stabilisation effect water has on polypeptides, the immobilisation of water around biomolecules and the observation of the electrostrictive effects where the water structure is severely disrupted by the presence of the peptides [11, 15, 34].

Although a single amino acid in water may not be necessarily an exact replica of a particular biological phenomenon, small biologically representative molecules in solution enable the test of experimental techniques, the validation of models, as well as investigations into the smaller-scale interactions that ultimately determine how larger systems behave. Small 'model' systems are particularly important if the use of *ab initio* calculations is required as large systems are computationally intractable. Thus to gain an understanding of larger biomolecular systems it is important to consider the smaller constituents as well as the larger whole.

Each amino acid is characterised by its hydrophobicity and hydrophilicity. The hydrophobic residues are usually placed within the middle of the protein, with hydrophilic side-chains exposed to bond to the aqueous environment. The human

body uses over 50,000 unique protein structures, each with unique structural purposes from bones, cell membranes to enzymes that act as catalysts. The property of each protein is determined by the arrangement of its constituent amino acids, of which there are only twenty. Each amino acids' functional groups, hydrophobicity and size is paramount in determining the final functionality of the protein that the amino acids assemble into. For example, in cell membranes the amino acids are arranged so that the exposed outer layer is hydrophobic, encouraging the formation of a lipid bilayer around the membrane. Each amino acid has unique properties, such as the flexibility of glycine or the cyclical ring of the proline side-chain.

Amino acids all contain two groups; an amine, -NH2, and a carboxyl, -COOH. Proteins are formed when amino acids link *via* the condensation reaction, where these two groups join producing a water molecule as they combine, and as more amino acids link up in a chain, a protein is formed. The link between these molecules, the peptide bond (-C(=O)N(H)-), is the basic repeating unit of any protein, with the specific sequence of amino acids down the length of the repeating unit forming the primary structure that defines a specific protein. Additional structural properties in proteins occur when interactions occur across the protein backbone. Each amino acid has a functional group that stands out from the backbone when the peptide bonds are formed. These functional groups can interact, and in doing so produce secondary structures such as alpha helices, where hydrogen bonds form between along the backbone forming a right handed spiral, and beta sheets which form when parallel β-strands of proteins (polypeptides of 3 to 10 members) interact with hydrogen bonding.

The type of secondary structures that form is directly related to the constituent amino acids. For example, beta sheets commonly contain amino acids with large aromatic residues (e.g. Tyr, Phe and Trp) and branched amino acids, such as Thr, Val and Ile, at the centre of the sheet, while at the edges of the sheet different residues may be prevalent such as proline. Dipole moments in parallel sheets may also influence the type of residue at the caps of the structure. It has been suggested that this may exist to discourage beta-sheet association which would lead to aggregation and amyloid formation, which occurs in brain diseases such as Alzheimers[35]. Alpha helices are a key functional role in providing stability in DNA binding motifs, they are also the most common protein structure

to cross biological membranes as the peptide backbone can be orientated to ensure that the sidechains are all hydrophobic. As secondary structures such as these depend on the functional groups, hydrogen bonding between amino acids, and the orientation of the peptide backbone, understanding these properties for smaller groups, such as a single amino acid, or large peptides is important for understanding the role biologically, and provides the main motivation for this study.

1.1.1 Glutamic acid

One amino acid that has sparked considerable interest in recent years is glutamic acid. It is one of the 20 proteinogenic amino acids, although it is made within the human body and thus is not one of the essential amino acids. *L*-glutamic acid ($C_5H_9NO_4$) is contained in many proteins and is considered a possible avenue towards the treatment of various neurological disorders such as Parkinson's disease, epilepsy and muscular dystrophy [36, 37]. It is also thought that glutamic acid enables the transmission of potassium across the blood/brain membrane. Glutamate has also been linked to strokes, mental retardation and Alzheimer's disease [38]

The salt, glutamate (shown in figure 1.1), is an important neurotransmitter and is important for memory and learning. It is also essential in the metabolism of fats and sugars. However, the use of it as a flavour enhancer in the form of monosodium-glutamate (MSG) has sparked speculation in the media over health concerns [39]. There is still a debate over whether directly ingested 'free' glutamate, i.e. not a constituent of a protein, may cause or exacerbate various neuroendocrine disorders, brain lesions and possibly lead to learning disabilities [40, 41, 42, 43]. Although there is no conclusive evidence of the effect of digested glutamate, understanding by which glutamate interacts with other parts of the body may provide the first avenue of approach in explaining these disorders.

Glutamate has an overall negative charge, and when residing in a protein provides a site that bonds to positively-charged molecules. The three charged sites on the glutamate zwitterion, one positive, and two negatively charged, provide multiple interaction sites when the amino acid is in isolation. As the presence of glutamate 'free' glutamate is attributed to the dietary concerns associated

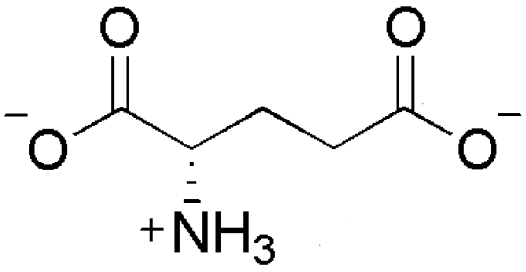


Figure 1.1: Schematic of glutamate.

with the salt, it is important to understand the amino acid in isolation, in an environment similar to that in the human body, namely an aqueous solution.

Glutamate molecules in aqueous solution have recently been studied by Leenders *et al.* [44] and Daub *et al.* [45]. Leenders *et al.* considered classical and *ab initio* simulations of both glutamate and glutamic acid molecules in solution [44]. They found that each of the two carboxyl groups on a single molecule interacted differently with the surrounding water molecules, in particular one carboxyl had a greater hydration number than the other. Daub *et al.* also considered simulations of aqueous glutamate, but at different concentrations [45]. Again they identified key differences in the behaviour of the charged groups. No analysis of the glutamate-glutamate interactions were considered in either of these studies.

In 2006 McLain *et al.* published results on the neutron diffraction and empirical structure refinement (EPSR) modelling of glutamate, in solution with sodium ions and water at a concentration of 1:1:29 (glu:sodium:water) [3]. This study did not provide individual occupation numbers for water molecules around the two different species of carboxyl. However, they did find that the distribution of water molecules around the two carboxylates were similar in the EPSR study. They also noted that the water structure was 'severely disrupted', with the average number of hydrogen-bonds per water molecule reducing from 1.8 in pure water to 1.4 in the glutamate solution. McLain *et al.* also observed that the three charged groups on the water molecule, the amine, NH_3^+ and the two carboxylates, CO_2^- were responsible for the reduction in water structure as the water-water bonds appeared to be substituted for glutamate-water bonds. Leenders *et al.* and Daub *et al.* also observed this change in water structure, although the effect was less pronounced in the lower-concentration systems [44, 45]. The quantum

mechanical simulations by Leenders *et al.* also found that two oxygens atoms in the same carboxyl group of the glutamate molecule sometimes behave differently, with different proportions of solvation occurring. None of this work considered the interactions between two or more glutamate molecules.

Studies of amino acids such as L-alanine (+NH3-C2H4-COO-), provide comparable results for studying glutamic acid. Alanine has an overall neutral charge, but shares the common bonding sites of the -NH_3^+ and the -CO_2^- that glutamate has. The study by Degtyarenko *et al.* found that the two oxygen atoms in the carboxyl group behaved differently in an aqueous environment, with one forming 2.34 bonds with water, and the other just 1.88 bonds [46]. This could affect how hydration shells form around a protein with this carboxyl group exposed, and if the effect is also found in alanine-alanine interactions then it could have significant implications to the required orientation of alanine in a protein to provide stable secondary structures. If similar effects are found in glutamate then these would have similar implications to its functionality in biological systems. Degtyarenko *et al.* also found that in the gas phase alanine molecules existed in their neutral non-ionic form, and they suggest that stable zwitterions only exist when hydrogen bonds can form.

The stability of amino acids in their zwitterionic state depends on the surrounding conditions stabilising the ion. Theoretical calculations of glycine suggest that it is not stable as a zwitterion in gas phase, but requires two water molecules to maintain the ionic state [47]. Other studies have suggested glycine requires 5 water molecules [48]. Further studies have shown that the presence of water around glycine can lead to a closed-ring structure of glycine forming. The only negatively charged side-chain amino acid other than glutamic acid is aspartic acid. Studies on the stabilising effect of aspartic acid on collagen show that the position of the asp amino acid in the protein can either significantly stabilise or destabilise the collagen as it increases the chain flexibility and changes inter-molecular hydrogen bonding patterns [49]. Similar effects may be found in systems containing glutamate, as the studies by Daub and Leenders suggest significant perturbation of the water structure.

As the simulations of aqueous glutamate also contain ions to neutralise the charge, it is important to consider the effect of ions on the system. As with the previous studies on glutamate we have chosen to consider a solution of

glutamate with water and sodium ions [45, 44, 3]. Ion interactions with water have undergone extensive study, however, less well understood on an atomistic scale is their effects on amino acids. Ions, such as sodium have a key role to play in aqueous, protein environments. The type of ion in a system changes ion adsorption and permeability of red blood cells. Ions also influence the activation energy that is required to 'cut' DNA with enzymes, and the type of ion present changes the required energy level. Boström *et al.* argue that many aspects of biological mechanics that have been attributed to the charge of ions and their ability to effect the surrounding species does not consider that the electrostatic potential is strongly screened at the concentrations often found in biology, as much of the theory relies on water-ion based simulations alone[50], in particular at membrane interfaces. Clearly the inclusion of sodium ions in a glutamate solution can provide useful input into this debate, particularly if compared with sodium ions in pure water, where there is a clear perturbation of the water structure, resulting in hydration shell around sodium ions of at least 5 members [51].

1.1.2 The tripeptide glycine-proline-glutamate (GPE)

As glutamic acid is an essential component in neurotransmission the focus in the 'bottom-up' approach of molecular modelling followed here suggested that the next logical course would be to consider the tripeptide glycine-proline-glutamate (gly-pro-glu or GPE). This tripeptide has been studied extensively from a neurological viewpoint as it appears to have neuroprotective properties. In particular, it may provide an avenue towards treatment of Parkinson's disease [52, 53] and Huntington's disease [54].

GPE is considered the key-part of the insulin-like growth factor-1 (IGF-1), which helps with long-term memory function, particularly after brain injury. GPE is easily cleaved from the N-terminal end of IGF-1 and has been studied for its neuroprotection. The methods *via* which IGF-1 and GPE aid the brain have not been fully elucidated.

Truncated IGF-1 was identified in fetal and adult brain tissue in 1986, but consisted of only 29 amino acids, without the final GPE tripeptide [55, 56]. The full IGF-1 sequence was not identified until 1992 [57], and then subsequently the intact GPE tripeptide was also found in the brain. The first work to establish the biological role of IGF-1 and GPE was carried out by Sara *et al.* in 1993

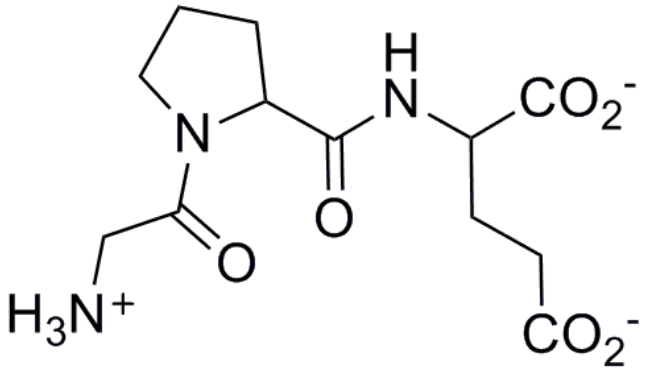


Figure 1.2: Schematic of the GPE molecule.

where they identified the importance of both in the central nervous system [58]. In particular, they suggest that due to the form of GPE, the tripeptide may interact with glutamate receptors and thereby play a role in neurotransmission as glutamate does. In 2006 de Diego *et al.* found that the neuronal protection was not due to GPE's affinity to glutamate receptors.

This summary, although not exhaustive, of the work on GPE, highlights that despite the fact that the importance of GPE in neuronal activity has been clearly demonstrated, the method *via* which this occurs is unclear. Our work aims to take a molecular view of this system. By understanding the molecular mechanisms involved in GPE interactions, the understanding of GPE's role in the body may be developed further in the future.

To our knowledge the only molecular simulation work currently published on the GPE molecule considers frequency fluctuations by studying the ultrafast dynamics of the molecule [59].

1.1.3 Water and sodium ions

In both the glutamate and GPE simulations water is included as we are interested in the solution structure. The GPE tripeptide and glutamate both have an overall charge of -1 e per molecule, and each system contains an equal number of sodium counter-ions in the solution as glutamate/GPE molecules, thereby neutralising the overall charge of the system. Sodium ions are particularly relevant in the glutamate system as monosodium glutamate is the glutamate salt with the addition of a single sodium ion. In addition McLain *et al.* have

provided experimental data from neutron diffraction of glutamate-sodium-water solutions, which has enabled direct comparison with simulations in this body of work.

Both water and sodium ions in solution have been studied extensively (for example see refs [6, 12, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70]). These studies have included the effect of sodium ions on the surrounding water molecules, including the relative bond lengths, and the dynamics of pure water and sodium-hydroxide solutions. The aim of this work is not to investigate water and sodium ion solution behaviour explicitly, as this has already been done extensively. Instead this work will focus on how the water and sodium ions behave because of the surrounding GPE and glutamate molecules. As mentioned above, there is some debate over the current understanding of how ions behave around proteins, and whether current models are relevant as they are mainly based on solutions of ions in water, rather than involving other biological molecules.

As the classical simulations considered in this work use the CHARMM22 classical forcefield [71] to parametrise the large molecules, which is designed to work with a modified version of the TIP3P water potential [72], we have provided a brief investigation into the effect of three different water models. The TIP3P water model is known to under-estimate the tetrahedral nature of pure water [73]. We consider two other important models for water that are representative of the types of water potential available for classical study of Molecular Dynamics systems. This is discussed further in chapter 2.

1.2 Thesis Outline

Chapter 2 provides an overview of the methods used within this study including the simulation and analysis techniques.

Chapter 3 is concerned with the structure of aqueous glutamate in solution, using both classical and quantum mechanical simulations.

Chapter 4 considers two further approaches to the study of the glutamate system. The first part examines the dynamics of the system and the second part considers the application of parallel tempering, an enhanced sampling technique.

Chapter 5 takes the next step in the 'bottom-up' approach by considering the structure of the tripeptide GPE from molecular simulations.

Finally, chapter 6 summarises the key observations from this body of work and suggests further avenues of research to be considered.

Chapter 2

Simulation and analysis methods

2.1 Introduction

Very few non-trivial problems in statistical mechanics are exactly soluble. Often, the more complicated a problem, the more interesting an exact solution would be. However, exact results or even appropriate approximations quickly become intractable as the complexity of the problem increases. Computer simulations offer an alternative avenue of approach, providing a solution to an approximation of the real problem. The solutions can be tested by comparison with experiment, and probed to test theories that are not accessible with current experimental techniques.

The relationship between a real system, simulation and experiment is complex, as shown by figure 2.1. Initially the simulation's approximations can be validated by comparison with the experiment. Once validated, simulations can provide important insight and information that experiments cannot extract. This then provides extra information for interpreting experimental results, and the most favourable outcome is when simulation can provide targets for future experiments and ideas for new technology for experiment.

In the study of complex liquids the relationship between experiment and computer simulation has become increasingly important over recent years. Simulation provides avenues to characterise both microscopic and macroscopic properties of a system, and can investigate properties on an atomic level not easily obtainable with current experimental techniques.

Molecular systems move and fluctuate, and indeed in the case of biological

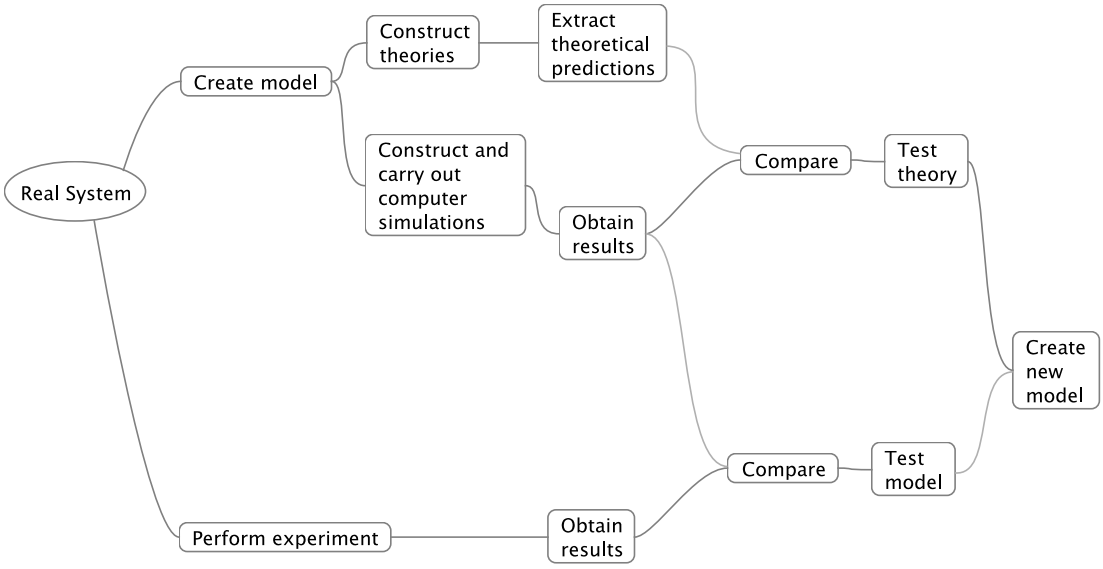


Figure 2.1: Relationship between computer simulation and experiment. This has been adapted from a similar diagram in Computer Simulation of Liquids by M.P. Allen and D. J. Tildesley [1].

macromolecules these movements can be intrinsic to their biological function. Thus, for simulations to gain meaningful results thousands or even millions of different geometries need to be sampled. To probe this geometrical landscape one may use either the stochastic or deterministic approach. The first is most easily achieved by Monte Carlo studies [1], whereby the energy surface is randomly probed by changing the geometry of the system. The alternative, deterministic approach, using Molecular Dynamics (MD) simulates a system over time. Thus, the full thermodynamic properties of a system can be obtained over the progress of a simulation and time-dependent properties can be determined. Due to the nature of many liquids, MD simulations can provide a thorough sampling of the local energy landscape as energy barriers are often small [1]. In addition, the high density of molecules in a liquid state decreases the efficiency of Monte Carlo techniques as a high proportion of the selected random moves will result in overlapping molecules, and will therefore be rejected. In the following study we have chosen to use MD simulations. This enables structural and dynamical properties of the systems to be evaluated.

Conclusions drawn from simulation should consider the propensity of a system to remain in a particular state, requiring an understanding of the entire energy, or

conformational, landscape available to a system. Simulations of simple systems can quickly cover the available conformational landscape, but simulations of molecular systems including hydrogen bonds are restricted in their movement and take far longer to over come energy barriers between possible configurations, resulting in poor sampling of the energy landscape. This can be overcome by considering much larger systems or simulating systems over a longer time-scale, but both approaches are restricted by computational power and time, which although increasing rapidly still limits the simulation of complex liquid systems. To overcome this limitation the enhanced sampling technique 'Parallel Tempering' is considered in the study of aqueous glutamate.

This chapter is split into two sections. The first covers the principles of Molecular Dynamics simulation techniques including classical and first-principles approaches and the packages used. The second section explains the analysis methods used to extract meaningful information from the simulations studied in this work.

2.2 Simulation techniques

The principle simulation method used in this work is Molecular Dynamics (MD), as it is fast and moderately efficient at simulation of aqueous solutions [1, 74] and provides a deterministic trajectory of the atoms through time. In particular, MD is generally quicker at simulating molecular systems as it does not need to reject unphysical configurations as Monte Carlo methods may do, but only pursues configurations determined by the properties of the atoms being simulated. The collisions inherent in liquid simulations aid MD in overcoming energetic barriers between conformations. In addition, MD can provide information on the time-evolution of molecules that is not available from Monte Carlo techniques, thereby enabling understanding of the mechanics of changes in structure.

Simulations of biological phenomena, protein fragments and entire proteins are often concerned with inter-species structures - how molecules behave around each other, especially with respect to an initial set of conditions, and how water and biologically important molecules interact. In this instance the time-evolution provided by MD algorithms is suitable for simulating such systems. Simulations are limited in size and length; for large proteins only hundreds of picoseconds of

trajectory data may be available even with the use of modern supercomputers.

In MD simulations the *ergodic hypothesis*, the assumption that the time average is equivalent to the ensemble average, justifies the use of MD as a time-evolution of a system that is representative of the full set of states available to the system. This means that, according to statistical physics, a macroscopic observable, A, such as energy, can be taken as the long-time average of the instantaneously observable property $A(t)$ that is the weighted average of the microscopic observable a_i;

$$A = \langle a \rangle = \frac{1}{N} \sum_{i}^{N} p_i a_i, \tag{2.1}$$

where p_i is the probability of being in state i over the entire ensemble of N microstates available. This probability manifests into the amount of time a system spends in a particular state, either in simulation or computation, and thus, should consider the infinite time-limit to sample all possible states. In reality simulations are limited to a finite time and, thus, the question of whether a sufficient proportion of the states has been sampled becomes an issue. If a simulation is too short, an insufficient sample of the possible conformations of a system are obtained, resulting in misleading results. This can be overcome by enhanced sampling *via* tools such as parallel tempering, which are discussed further below.

There are two main approaches to MD simulations, the first employs an empirical potential energy function, or force-field, to describe the interactions of the system. This classical approach is limited by the accuracy of the potentials, which are often developed to reproduce one or two key properties of a system and, thus, may not provide a complete characterisation of all the properties of the system under consideration. Thus, the second approach, *ab initio*, or first-principles, MD utilises quantum mechanics to provide a simulation of a system *Ab initio* MD is based on the solution of the time-independent Schrödinger equation. First principles methods are a variant of *ab initio* MD that is more efficient, although it is not strictly speaking a truly *ab initio* approach. Truly *ab initio* methods use an initial estimate of the molecular orbitals, constructed from a plane-wave basis set, that is then explicitly solved for the minimum electronic state, thereby correcting the 'guess'. Instead of using full molecular orbitals,

the first-principles approach characterises the electronic state by the use of an exchange/correlation functional and associated pseudo-potentials, and thus the accuracy of the simulation depends on the accuracy of the functionals used. Both *ab initio* and first principles MD are also limited by their computational requirements as they scale as N_e^3 for N_e electrons. The methods of classical and first-principles MD are visited in more detail later.

The following sections outline the basic principles and methods of both classical and first-principles Molecular Dynamics. They are not intended to be a complete and thorough treatment of the theory or the literature available on the subject. For further information on the methodology refer to references [1, 74, 75, 76].

2.2.1 Basic principles of Molecular Dynamics

Molecular Dynamics (MD) simulates molecules on an atomistic basis, with each atom having an initial position and velocity. The potential energy and resulting forces on each atom are then computed, allowing the positions to be updated using Newton's equations of motion. This process is then repeated until the desired number of steps has been completed, as shown in figure 2.2.

The initialisation of the system, choosing appropriate positions and velocities, needs to be performed such that they are compatible with the target simulation. The force calculation is usually the most time-consuming step of an MD simulation. Every single atom, or particle, is considered, and in each case the force on each particle may have contributions from all other particles. This results in $N \times (N-1)/2$ pairs to be considered, resulting in a computational time scaling with N^2. In modern MD simulations, this is often reduced by using neighbour lists, whereby non-bonded interactions are handled by the use of a Verlet neighbour list. The list contains references to all secondary atoms within a cut-off distance of all primary atoms, and only these interactions are included in the long-range force calculations, avoiding computation of all pairs of atoms throughout the entire volume of the simulation. Finally the integration of the equations of motion, which are dependent on the forces calculated, governs the dynamical behaviour of the particles. Appropriate use of algorithms and sensible choice of time-step size are used to minimise accumulative error in the simulation while keeping the real computer time to a manageable level.

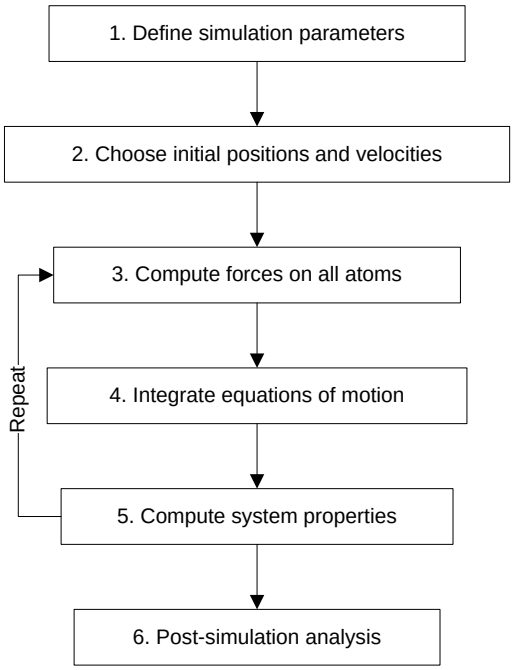


Figure 2.2: Generic algorithm for MD.

Periodic boundary conditions

In an ideal world an infinite system could be simulated such that we could observe the real world in the detail available from computer simulation. One of the main limitations faced by computer simulators is the limitation of computational power, which, while increasing rapidly, are far from being able to simulate every molecule in the human body simultaneously. Thus, small numbers of atoms are considered as representative sub-systems. This can result in artificial behaviour with finite size effects and surface-interactions at the edge of a box affecting the overall system behaviour.

One of the methods to overcome the limitations of a finite system and surface effects is to use periodic boundary conditions where molecules leaving one side of the box will re-enter at the opposite side. In addition 'periodic images' are used in the calculation of inter-particle forces. In this instance the minimum-image convention is used, in that particles only interact with the nearest 'image' of another particle. For instance, if we are interested in the interaction between atom A and atom B, convention would suggest that this results in the use of the inter-atomic distance $\mathbf{r}_{AB}$ where the radial distance is from within the central simulation box and does not account for periodic boundary conditions. The

shortest distance between the atoms may not be through the visualised box, but may be through one of the boundaries, utilising the periodic nature of the box. In this case the periodic boundary distance can be used, otherwise known as the minimum-image distance.

Ensembles

As MD simulations are on a closed system they conserve the number of atoms, N. They also rely on the basic premise of statistical physics that information at the microscopic level (e.g. the positions and velocities), can reproduce macroscopic properties such as pressure and energy (see the *ergodic* hypothesis above). This requires that there are a constant number of atoms, N, within the system. In an equilibrated system this often results in a system that is close to maintaining a constant volume. This results in the 'natural' ensemble for MD, the microcanonical ensemble, where N, volume V and energy E are conserved during the simulation. MD simulations require that these basic 'state' variables are conserved during the integration of the equations of motion. However it is often more useful to simulate other ensembles, such as NVT or NPT where temperature, T and pressure P are used as the conserved state variables for the system. Performing experiments with constant energy is virtually unheard of, thus, NVT and NPT ensembles enable more realistic comparison.

In the following work the Nosé-Hoover canonical (NVT) thermostat and the Berendsen isobaric-isothermal (NPT) thermostat/barostat are used [77, 78]. In the Berendsen thermostat constant temperature is achieved by scaling the velocities at each time-step. The scale factor is given by

$$\left[1 + \frac{\delta t}{\tau_T}\left(\frac{T}{T_0} - 1\right)\right]^{\frac{1}{2}}, \tag{2.2}$$

where the temperature T is scaled towards the target temperature T_0 using a relaxation of 'rise-time' constant τ_T, and δt is the simulation time-step size. The larger the value of τ_T, the weaker the coupling. Berendsen found that for static structure and dynamics to be adequately reproduced in a simulation of water that a relaxation time of 0.4 ps was needed [78]. In addition to constant temperature the system is kept at constant pressure by re-scaling the size of the simulation box resulting a Berendsen thermostat and barostat.

The Nosé-Hoover thermostat [77] builds on the work by Nosé in 1984. A canonical ensemble, constant NVT, is achieved by modifying Newton's equations of motion to include a friction term, $\chi(t)v(t)$, where the friction coefficient χ, depends on the difference between the instantaneous and desired kinetic energy of the system. This enables correction of the temperature towards the desired temperature. The time-evolution of the 'correction' is governed by a time-constant normally in the range of 0.5-2.0 ps.

Nosé-Hoover Chains

One limitation of the Nosé-Hoover thermostat is that, in exceptional circumstances, a canonical ensemble is not generated, i.e. the thermostat is not strictly *ergodic*, as the desired distribution is not obtained. This is normally due to a non-zero velocity of the centre of mass of the system. (For more details on this the reader should refer to Frenkel and Smit [74].) To overcome this Martyna *et al.* suggest the use of Nosé-Hoover chains where the standard thermostat is coupled to one or more other thermostats [79]. Again a suitable relaxation time needs to be provided.

Integration of the equations of motion

Molecular Dynamics works on the principle of moving a system forward in time in short incremental time-steps. In the simplest classical case Newtonian mechanics is used to evaluate $\mathbf{F} = m\mathbf{a}$ and propagate the system forward in time appropriately. Firstly all the potentials within the system are calculated, then this is used to find the forces within the system which can then be used to move each particle using simple mechanics. Solving a set of differential equations for a Newtonian system is thus, relatively simple.

Once the forces within the system have been found a *finite difference* method is used to obtain the positions and velocities at time $t+\delta t$. This method uses the position and velocities of a particle at time t to derive a time-evolved equation for the positions and velocities at time $t + \delta t$. One such scheme is the Verlet Leapfrog algorithm [80], discussed later, which is a simple and reversible finite difference method.

The basic Verlet algorithm calculates the position in the next time step, *time* $= t+\delta t$, from the positions at the previous *and* the current step, thereby reducing

the errors introduced in using just using the current. The algorithm is derived from the two Taylor expansions:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{\mathbf{a}(t)\delta t^2}{2} + \frac{\mathbf{b}(t)\delta t^3}{6} + \mathcal{O}(\delta t^4), \tag{2.3}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{\mathbf{a}(t)\delta t^2}{2} - \frac{\mathbf{b}(t)\delta t^3}{6} + \mathcal{O}(\delta t^4). \tag{2.4}$$

, were $\mathbf{v}$ (velocity), $\mathbf{a}$ (acceleration) and $\mathbf{b}$ are the second, third and fourth order derivatives of the distance $\mathbf{r}$. Addition of these two equations results in the equation to advance position over a specified time, δt;

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2\mathbf{a}(t) + \mathcal{O}(\delta t^4). \tag{2.5}$$

Using two time-steps gives an error of order δt^4. In addition, the method relies on two time-steps t, and $t - \delta t$, and thus, the first time-step is subject to greater error than subsequent time-steps.

Velocities, although not strictly needed for this integration scheme, are often useful in simulated systems. In the Verlet scheme they are obtained by the approximation:

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t}. \tag{2.6}$$

As this uses positions two time-steps apart the error in the velocity is of order δt^2.

To overcome these problems the Leapfrog Verlet (or Velocity Verlet) expands on the basic Verlet algorithm to incorporate velocity, thereby removing the first time-step error and providing a better method of handling the velocities. In this method we Taylor expand $\mathbf{r}(t+\delta t)$ and $\mathbf{v}(t+\delta t)$ and using the same method as in the basic Verlet algorithm we obtain

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{\mathbf{a}(t)(\delta t)^2}{2}, \tag{2.7}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{\mathbf{a}(t) + \mathbf{a}(t+\delta t)}{2}\delta t. \tag{2.8}$$

The implementation of the algorithm requires the position and force (related to $\mathbf{a}(t)$) at time t while the velocities are half a time-step behind. This results in

a two-step algorithm where the velocities are advanced a half time-step:

$$\mathbf{v}(t + \frac{1}{2}\delta t) = \mathbf{v}(t) + \frac{1}{2}\delta t\mathbf{a}(t). \tag{2.9}$$

This is followed by updating the positions of the particles using the half time-step velocities. As the velocities at time-step t may be required, the discrete time-step $\mathbf{v}(t)$, rather than the half time-step velocity, can be obtained from the average of the two velocities a half time-step either side of the desired time.

This algorithm is fast and efficient and only requires storage of one set of positions and velocities, although obtaining the velocity requires storage of the previous time-step velocity as well. Truncation of the Taylor expansion compounds the errors of size $\mathcal{O}(\delta t^4)$ in the positions and of size $\mathcal{O}(\delta t^2)$ in the velocities, at each time-step. It is therefore essential for time-steps to be small enough to minimise this.

As this method (along with all Verlet algorithms) is time-reversible, little energy drift is experienced when using the Verlet leapfrog algoriths. This coupled with the simplicity of the Verlet algorithms, have resulted in it becoming one of the most commonly used integration schemes in MD simulation. The Velocity Verlet offers the added benefit of not using positions that are two time-steps apart as the standard Verlet scheme does. This scheme is used in both the classical and first-principles simulation schemes of the MD codes we have used in this study.

2.2.2 Classical Molecular Dynamics

Ideally simulations of molecules would involve a full quantum mechanical description of each of the electrons and nuclei in a system, however, this is not currently possible. An alternative is to simulate particles using various quantum mechanical approximations, however, this approach has various limitations. Firstly, only small systems can be used and even those require vast computational resources. Secondly, the systems require very small time-steps which are slow to calculate and thus, are limited to very short time-scales. A quantum mechanical approach is discussed in section 2.2.3. Itt is necessary to also consider simulations that enable larger systems and longer simulation times. Classical approximations remove these limitations and, in the correct circumstances, can be accurate enough to provide useful information to provide a realistic simulation to compare

to experiment.

The simplest classical approximation is to describe atoms as hard spheres, the resulting molecules as rigid or semi-rigid structures comprised of the hard spheres and with fixed bond and torsion angles. Finally, simple inter-atomic potentials can be used to replace chemical bonds in a full system description comprised of multiple molecules.

This simple approach assumes that bond vibrations are of a small amplitude and can, thus, be ignored. For large molecules this results in a reduction of the accuracy of the simulations as it restricts the degrees of freedom and reduces the total number of conformations that the system can achieve. The alternative is to enable some flexibility within the bond angles, while keeping a fixed bond length. In addition, some degrees of freedom are left completely un-restricted enabling long chains to rotate a full 360° in one plane. However, the models quickly become more complicated resulting in more complex potential energies, forces and torques.

The potential energy is divided into many terms, describing the various effects within a multi-particle system. Each term in the potential depends on the spatial co-ordinates of an individual atom.

$$V = \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j>j} v_2(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + ... \qquad (2.10)$$

The first term is the sum of the effect of an external field on each particle, i. The other terms give the particle interactions; higher order interactions are less significant and much more computationally expensive to compute and, thus, are usually left out in most classical simulations. The potential can then be used to solve the classical equations of motion.

In classical MD a 'forcefield' is used to describe the potential that governs particle motions. In the study of aqueous glutamate and GPE solutions, the CHARMM22 potential was used to describe the glutamate and GPE molecules [71]. The water molecules were characterised by three different potentials, F3C, SPC/E and TIP3P [81, 82, 72]. CHARMM22 is an all-atom force-field, derived from first-principle calculations where model components interacted with water. CHARMM27, built on the work for CHARMM22, so that it could be used for RNA, DNA and lipids. Both CHARMM parametrisations are

available from the MacKerrell website [83]. CHARMM22 is a standard forcefield used in the simulation of protein fragments. It is parametrised to work with a modified version of the TIP3P water potential developed by Jorgensen *et al.* [72] and so this is the first water potential considered. However, TIP3P has been shown to poorly reproduce many key features of water, in particular the tetrahedral structure of pure water [73].

Nutt and Smith [84] studied aqueous *N*-methlyacetimide solutions using the CHARMM22 empirical potential with TIP3P, TIP4P and TIP5P water potentials [72, 85, 86]. They concluded that, although there were differences in the binding energies and 'minimum energy geometries' , the other potentials are suitable for use with CHARMM22 [84]. Thus, given the limitations of the TIP3P potential we consider two further models. Firstly SPC/E, a rigid model like TIP3P, that reproduces key features of water at room temperature. Secondly we consider the flexible potential, F3C, which uses the initial geometry of SPC/E but enables flexibility in the angles. A summary of the different geometries for the three water models is given in table 2.1.

	TIP3P	SPC/E	F3C
Angle	104.52°	109.47°	109.47°
O-H bond length	0.9572Å	1.00 Å	1.00 Å

Table 2.1: Definition of the geometric parameters for the three water models; TIP3P, SPC/E and F3C. The geometry of the F3C model is the equilibrium values only.

For each atom the sum of the potential terms that result in the force used in the integration method include pairwise chemical bond terms, angular constraints and non-bonded interactions in the form of *van der Waals* terms. In addition larger molecules, such as glutamate and GPE are subjected to the intra-molecular dihedral torsions, and the inter-molecular pair-body, three-body, four-body and many-body covalent forces.

Harmonic bond-potentials for atoms α and β at a separation distance $r_{\alpha\beta}$ have the form

$$U(r_{\alpha\beta}) = \frac{1}{2}k(r_{\alpha\beta} - r_0)^2, \tag{2.11}$$

where r_0 is the equilibrium bond separation, and k is a constant. This type of interaction describes an explicit bond between specified atoms. The angle

potentials for F3C and CHARMM22 take a similar form with a harmonic potential for the angle between three atoms, $\alpha\beta\gamma$ constrained to the equilibrium angle $\theta_{\alpha\beta\gamma}$ by:

$$U(\theta_{\alpha\beta\gamma}) = \frac{1}{2}k(\theta_{\alpha\beta\gamma} - \theta_0)^2. \tag{2.12}$$

Dihedral angles constrain a group of four atoms to a particular orientation, where atoms A, B and C reside in one plane and atoms B, C and D reside in a different plane. The dihedral constraint reduces the freedom of the planes with respect to each other to describe the interaction arising from torsional forces arising in molecules. The CHARMM22 forcefield provides values for a dihedral constraint of the form:

$$U(\phi_{\alpha\beta\gamma\delta}) = A(1 + cos(m\phi_{\alpha\beta\gamma\delta} - \delta)), \tag{2.13}$$

where A is a constant of the potential, $\phi_{\alpha\beta\gamma\delta}$ is the angle between the two planes and δ is the equilibrium value desired. The values for A and δ are provided in the CHARMM22 database. As classical forcefields are an approximation, improper dihedrals are also included. These do not correspond to any chemical feature but enable the restriction of the molecule further so that it maintains correct molecular behaviour. This allows the constraint of four atoms *via* the same potential as used for dihedral angles.

The first inter-molecular interaction is a *van der Waals* interaction in the form of a Lennard-Jones potential:

$$U(r_{\alpha\beta}) = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}}\right)^{6}\right], \tag{2.14}$$

where $\epsilon_{\alpha\beta}$ is the depth of the potential well and $\sigma_{\alpha\beta}$ is the intermolecular distance at which the minimum occurs. This is a short-range pair-wise force that is parametrised for all three water models and for the CHARMM22 potentials. In each case $\epsilon_{\alpha\alpha}$ and $\sigma_{\alpha\alpha}$ are provided for a single type of atom interacting with another atom of the same species, and pair-wise forces are calculated for all pairs, avoiding $\alpha = \beta$. CHARMM22 uses the following combination rules to combine ϵ_α and ϵ_β to obtain $\epsilon_{\alpha\beta}$ and similarly for σ:

$$\epsilon_{\alpha\beta} = \sqrt{\epsilon_{\alpha\alpha}\epsilon_{\beta\beta}} \tag{2.15}$$

$$\sigma_{\alpha\beta} = \frac{1}{2}(\sigma_{\alpha\alpha} + \sigma_{\beta\beta}) \tag{2.16}$$

The Lennard-Jones interaction decays quickly with distance (as r^{-6}) and, thus, it is acceptable to only consider atoms within a certain distance of each other, thereby massively reducing the computation time.

The second inter-molecular interaction is the long-range electrostatic Coulomb interaction. Long range electrostatic interactions involves computing the interaction between all pairs of atoms, and cannot be subjected to a short-range cut-off as it is proportional to r^{-1}. This can be a very computationally intensive calculation, and thus the use of less CPU intensive approximations are commonly used. There are many methods available, including a direct calculation of the Coulomb sum, where the interaction potential for two charged ions, i and j, with charges q_i and q_j, is calculated as:

$$U(r_{\alpha\beta}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{\alpha\beta}} \tag{2.17}$$

The simplest way to calculate this is by the 'direct sum', where in a periodic system of N particles in a box of length L, the energy of the main cell is:

$$U = \frac{1}{4\pi\epsilon}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\mathbf{n}\in\mathbb{Z}^3}\frac{q_i q_j}{|\mathbf{r}_\alpha\beta + \mathbf{n}L|} \tag{2.18}$$

The sum over $\mathbf{n}$ is an infinite series as it runs over all of $\mathbb{Z}^3$, and therefore requires a cut-off, where it is assumed that contributions from larger $\mathbf{n}$ values can be ignored. This introduces an intrinsic error which decreases with the greater proportion of real space that is included in the calculation. In addition, as the force is long-range, the cut-off needs to be very large, still leaving a computationally intensive calculation. Creating an artificial sphere of interaction around a central atom by introducing a cut-off will not guarantee a neutral sphere of influence, and spurious 'charging' effects will be seen in the simulation. This can be removed by modifying the sphere to include neutral groups, but still results in poor physics and is only useful for simple structure optimisation calculations.

Due to the limitations of the direct sum, it is of little use for simulations. The method of choice for computing electrostatic quantities is the Ewald sum, based on splitting the infinite sum into two which can be computed faster (with accuracy fixed rather than dependent on the number of terms included) and quickly. The

derivation of the Ewald sum is detailed in below.

The Ewald summation method

The electrostatic interaction between charged particles adds potential energy to a system of the form:

$$U_c = \frac{1}{2}\sum_{i=0}^{N} q_\alpha \phi(r_i), \tag{2.19}$$

where particle i, with charge q, experiences the electrostatic potential $\phi(r_i)$ due to the other charged particles in the system;

$$\phi(r_i) = \sum_{j}^{N} {\sum_{\mathbf{n}}}' \frac{q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|}. \tag{2.20}$$

The prime indicates interactions with all particles, but excludes self-interaction. As the electrostatic potential does not converge quickly in a periodic system, the periodic images must be included. The box length, L is used to add on the appropriate distance of a charged particle in a periodic image, with the position of the periodic image defined by the lattice vectors $\mathbf{n}$.

The Ewald method modifies this summation over all pairs to contain a real space and a reciprocal space term, both of which converge quickly. This is achieved by spreading a'screening' charge over the system, neutralising all the point charges in the system. This is done by superimposing a spherical Gaussian screening charge or 'cloud' of the opposite charge, centred on the now-neutral ion. This results in the charge at a distance being due to just the fraction of the point charge that is not screened by the cloud at that distance. This is repeated for all ions, combining to form the *real* space term of the Ewald sum, with each term decreasing rapidly with distance and so can be viewed as a 'short-range' interaction. The *Reciprocal* space term (the 'screening' potential) is obtained by adding a second set of Gaussian charge 'clouds' over the first, of the opposite charge, neutralising the first set. For each set, the total sum is finite and easily summed in real/reciprocal space respectively. Combined, these two sums replace the point charge Coulomb interaction as the sum of two terms, but require a correction term to account for a Gaussian acting on its own site; a *self-energy correction*. The Gaussian charge clouds assume all particles interacting with

each other, however, particles should not interact with themselves, so a 'self'-correction term is added to remove correct for this added potential energy. Thus the interaction potential is re-written as:

$$U_c = U_{screening} - U_{self} + U_{short-range}. \tag{2.21}$$

To calculate the screening term we consider the electrostatic potential at a point r_i due to a charge distribution $\rho(r)$ that is constructed from the sum of periodic Gaussians:

$$\rho_{screen}(r) = \sum_{j=1}^{N} \sum_{\mathbf{n}} q_j (\alpha/\pi)^{\frac{3}{2}} \exp\left[-\alpha|\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|^2\right]. \tag{2.22}$$

Using Poisson's equation, $-\nabla^2_{screen}\phi(r) = 4\pi\rho_{screen}(r)$, and its Fourier transform, $k^2_{screen}\tilde{\phi}(k) = 4\pi\tilde{\rho}_{screen}(k)$, to transform the charge density obtains:

$$\begin{aligned}
\rho_{screen}(\mathbf{k}) &= \int_V d\mathbf{r} \exp(-i\mathbf{k}\cdot\mathbf{r})\rho_{screen}(r) \\
&= \int_V d\mathbf{r} \exp(-i\mathbf{k}\cdot\mathbf{r}) \sum_{j=1}^{N} \sum_{\mathbf{n}} q_j (\alpha/\pi)^{\frac{3}{2}} \exp\left[-\alpha|\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|^2\right] \\
&= \int_{allspace} d\mathbf{r} \exp(-i\mathbf{k}\cdot\mathbf{r}) \sum_{j=1}^{N} q_j (\alpha/\pi)^{\frac{3}{2}} \exp\left[-\alpha|\mathbf{r} - \mathbf{r}_j|^2\right] \\
&= \sum_{j=1}^{N} q_j \exp(-i\mathbf{k}\cdot\mathbf{r}) \int_{allspace} d\mathbf{r} \exp(-i\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}_j))(\alpha/\pi)^{\frac{3}{2}} \exp\left[-\alpha|\mathbf{r} - \mathbf{r}_j|^2\right] \\
&= \sum_{j=1}^{N} q_j \exp(-i\mathbf{k}\cdot\mathbf{r}) \exp(-k^2/4\alpha).
\end{aligned} \tag{2.23}$$

Inserting this into the Fourier transform of Poisson's equation gives

$$\tilde{\phi}_{screen}(k) = \frac{4\pi}{k^2} \sum_{j=1}^{N} q_j \exp(-i\mathbf{k}\cdot\mathbf{r}) \exp(-k^2/4\alpha). \tag{2.24}$$

As the derivation of the Ewald sum requires that this term converges, this means that the equation for $\tilde{\phi}$ is only valid for $\mathbf{k} \neq 0$. The Ewald method allows the sum to be set to 0 in the instance of $\mathbf{k} = 0$, which is consistent with the system

being embedded in an infinite dielectric constant.

Given the potential *phi*, then screening potential can then be calculated;

$$\begin{aligned} U_{screen} &= \frac{1}{2}\sum_{i=0}^{N} q_i \phi_{screen}(r_i) \\ &= \frac{1}{2}\sum_{\mathbf{k}\neq 0}\sum_{i,j=0}^{N} \frac{4\pi q_i q_j}{Vk^2}\exp(-i\mathbf{k}\cdot\mathbf{r}_j)\exp\left(-\frac{k^2}{4\alpha}\right). \end{aligned} \tag{2.25}$$

The self-correction term in the potential, U_{self}, removes the additional potential of self-interaction introduced by the screening potential:

$$U_{self} = \frac{1}{2}\sum_{i=0}^{N} q_i \phi_{self}(r_i). \tag{2.26}$$

ϕ_{self} is obtained by integrating Poisson's equation once more to obtain the electrostatic potential due to a spherically symmetric Gaussian charge, as used in the screening cloud:

$$\begin{aligned} \phi_{gauss}(r) &= 2q_i(\alpha/\pi)^{\frac{1}{2}}\int_0^r dr\exp(-\alpha r^2) \\ &= (q_i/r)erf(\sqrt{\alpha}r), \end{aligned} \tag{2.27}$$

where the error function, erf(x) is defined as

$$erf(x) = 2/\sqrt{\pi}\int_0^x \exp(-u^2)du = 2/\sqrt{\pi}(x - \frac{1}{3}x^3 + \frac{1}{10}x^5 - \frac{1}{42}x^7 + ...). \tag{2.28}$$

The Taylor expansion of the error function enables the calculation of ϕ at $r = 0$ (the self-interaction zone);

$$\phi_{gauss}(r) = 2q_i(\alpha/\pi)^{\frac{1}{2}}, \tag{2.29}$$

and the resulting potential becomes:

$$U_{self} = (\alpha/\pi)^{\frac{1}{2}}\sum_{i}^{N} q_i^2. \tag{2.30}$$

The final term in the potential, the short-range potential, is the sum of the

real space point charges along with the screening Gaussian distributions. Using the result obtained in equation 2.26, the electrostatic potential due to a point charge, that is surrounded by a Gaussian with equal and opposite net charge is,

$$\phi_{short-range}(r) = (q_i/r) - (q_i/r)erf(\sqrt{\alpha}r) \qquad = (q_i/r)erfc(\sqrt{\alpha}r), \tag{2.31}$$

where the complementary error function, $erfc(x)$, is defined as $erfc(x) = 1 - erf(x)$. The contribution due to the screened Coulomb potential is therefore:

$$U_{short-range} = \frac{1}{2}\sum_{i\neq j}^{N} q_i q_j \frac{erfc(\sqrt{\alpha}r_{ij})}{r_{ij}} \tag{2.32}$$

Adding the three potential terms together gives the electrostatic potential energy for the full Ewald sum of a periodic system of N ions:

$$\begin{aligned} U_c = & \frac{1}{2}\sum_{\mathbf{k}\neq 0}\sum_{i,j=0}^{N} \frac{4\pi q_i q_j}{Vk^2}\exp(-i\mathbf{k}\cdot\mathbf{r}_j)\exp\left(-\frac{k^2}{4\alpha}\right) \\ & - (\alpha/\pi)^{\frac{1}{2}}\sum_{i}^{N} q_i^2 \\ & + \frac{1}{2}\sum_{i\neq j}^{N} q_i q_j \frac{erfc(\sqrt{\alpha}r_{ij})}{r_{ij}} \end{aligned} \tag{2.33}$$

2.2.3 Car-Parrinello Molecular Dynamics

Classical simulations are limited in their ability to accurately reproduce key features of molecular systems. As classical potentials rely on the fitting of parameters to phenomena observed in a particular environment they are necessarily limited in their applicability. This is particularly clear from a review of 45 different water models by Guillot [87]. There is often no clear definition of how to define observed physical characteristics of molecular systems such as the hydrogen-bond, and bonds cannot break or form in classical simulations. This is overcome by using first principles simulation which explicitly treat the structure from an electronic perspective instead of relying on parameterisations.

Ab initio and first principles MD use methods based on the solution of Schrödinger's equation, thereby enabling the three-dimensional structure of a

molecular system to be determined from fundamental physics rather than classical approximations. These techniques do not require the inter-atomic potentials that are the weakness of classical simulations; instead, the simulation proceeds from accurate electronic structure calculations. The basic premise is to use Schrödinger's wave equation to describe the motions of the electrons and nuclei within atoms. This results in a substantial computational overhead, limiting the temporal and spatial scales of simulations. Using various approximations, in particular the Born-Oppenheimer approximation, where the motions of the electrons and nuclei are treated separately, simulations pico-seconds in length of small bio-molecular systems are possible with currently available computational resources. Once the forces acting on the nuclei due to the electronic structure are calculated then the molecular dynamics trajectory can be calculated.

In this work we have used Car-Parrinello *ab initio* molecular dynamics (CPAIMD) [88]. This method has the disadvantage of requiring a smaller time-step than other *ab initio* approaches; however, as complex liquid simulations by their very nature require a small time-step, the CPAIMD method is eminently suitable.

The CPAIMD method enables the explicit treatment of electrons while calculating the overall force of the system using density functional theory (DFT). It is based on the basic Velocity Verlet integration algorithm [89] which treats velocities explicitly, thereby enabling the integration time-step to be easily changed in the middle of a run. In addition, the CPAIMD method described by Tuckerman and Parrinello in 1994, and used in this work, uses the Nosé-Hoover chain technique [79] to enable better temperature control [89, 90, 91]. This results in the simulation being fully reversible and there is better conservation of energy, enabling larger time-steps to be used and therefore longer simulations.

DFT requires that the electronic potential can be described solely by the ground-state density, therefore if the ground-state density is known the energy can be calculated. The theory involves describing the electronic system by a functional (functions of functions), and for the purposes of describing the electronic potential, only the electronic ground-state is needed. DFT is based on the Hohenberg-Kohn theorem that the electrons can be described by the electronic density, but modified to replace the Hartree-Fock exchange energy with an exchange-correlation functional (see ref [92]). Theoretically, this should mean

that one functional could represent all systems. This 'perfect' functional has yet to be identified and therefore there are currently a multitude of density functional methods available.

The treatment of the electrons by DFT means that no assumptions about the bonding of a system need to be made. The method uses the Born-Oppenheimer approximation, in that it is assumed that the electronic ground state is an accurate and sufficient approximation of the electronic system, i.e. because electrons move much faster than nuclei, the electron motion is sufficient to characterise the motion of the entire particle. This approximation is valid as the mass of an electron is several orders of magnitude lighter than even a proton. CPAIMD is not a truly *ab initio* approach in that the slow nuclear motion is calculated with a classical potential. The problem of resolving the forces on an atom is split into two components; a quantum electronic calculation and a classical calculation for the nuclei.

In quantum mechanical simulations the wavefunctions used solve Schrödinger's equation, and are typically built up from a set of basis functions. These basis functions, or atomic orbitals, are combined linearly to construct the vectors of the wavefunctions. CPAIMD uses the LDA (local density approximation) Kohn-Sham equation to obtain the following Lagrangian:

$$\mathcal{L} = \sum_I \frac{1}{2} m_I \dot{\mathbf{r}}_I^2 + \sum_i \frac{1}{2} \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle - \langle \Psi_0 | \mathcal{H} | \Psi_0 \rangle + constraints, \tag{2.34}$$

where the first term contains the nuclear mass and separation, $\{\psi_i\}$ is the orbital, or basis functions for particle i, which can be combined into the wavefunction Ψ_0. The first two terms comprise the kinetic energy of the system and the third term is the electronic subsystem energy. The *constraints* term allows possible constraints to be imposed such as generalised orthonormality conditions. From this equation Euler-Lagrange equations are obtained for both the electronic orbitals and the nuclear positions. These are then solved to provide a force that is then used to integrate the positions and velocities of the particles.

The wavefunctions of electrons become more and more oscillatory in nature as $r \to 0$. This greatly increases the number of terms in the plane wave basis set, i.e. the number or orbitals, required to construct the wavefunction, and therefore the computational cost. This region is of little interest in simulations as the radial

distance, r, is very small. We therefore use *pseudopotential* approximations in the core region to enable a simpler wavefunction construction. This results in a less oscillatory small r region and consequently fewer terms in the Fourier series used to construct the wavefunction.

While many conditions apply to the accurate generation of pseudopotentials, they are an approximation and using this method means that the simulation moves further away from a truly *ab initio* approach. These approximations are required to make the problem tractable and provide simulations of useful size and time-scales.

The CPAIMD simulations carried out for this work use norm-conserving pseudopotentials [93]. This is where the phase-shift of the potentials is considered. A norm-conserving pseudopotential is set up such that

$$4\pi \int_0^{r_c} \mathrm{d}r \; r^2 R_\ell^2(r) = 4\pi \int_0^{r_c} \mathrm{d}r \; r^2 R_{\mathrm{ps},\ell}^2(r) \tag{2.35}$$

is obeyed, where $R_\ell(r)$ is the radial wavefunction of particle l and $R_{\mathrm{ps},\ell}(r)$ is the pseudopotential approximation. This means that within the region that the pseudopotential is used, $0 \rightarrow r_c$ (the cut-off radius), the phase-shift of the pseudopotential will be same as the real potential to first order in energy by making the wave-function identical to the original electronic wave-function outside the core.

As mentioned above there are a multitude of functionals available. The choice of functional is determined by their ability to reproduce experimental data, in a similar fashion to how classical empirical potentials are chosen (hence DFT and therefore CPAIMD are semi-empirical and not strictly speaking *ab initio* techniques). The simplest functionals involve approximating the electron density as a uniform gas, but this is poorly representative of molecular systems where electrons are not uniformly arranged. Instead gradient corrected methods are used, where the functional includes a gradient of electron density. One such method is the B-LYP, or Becke-Lee-Yang-Parr functional. The simulations carried out on glutamate using CPAIMD use the gradient-corrected B-LYP approximate density functional and the B-LYP plane wave basis set [94]. These have been shown to reproduce pure water effectively [95, 96].

For a thorough treatment of *ab initio* and Car-Parrinello methods the reader is referred to Marx and Hutter [92].

2.2.4 Further simulation techniques

In this work two further simulation techniques have been used in addition to classical and quantum mechnical simulation techniques described above. Firstly, energy minimisation using the congugate gradient method has been used to optimise the initial geometries of the systems under consideration and secondly, parallel tempering, an enhanced sampling technique, which overcomes one of the main limitations of molecular dynamics of liquids, namely the poor sampling of the total possible conformations of a system.

Energy minimisation

The energy minimization uses a classical force-field based conjugate gradient method to iteratively optimise the geometry of the N atoms in the system until the force convergence measure, $\Delta F = F_i.F_j < \Delta F_{threshold}$, where F_i and F_j are the forces at two different times, i and j and $\delta F_{threshold}$ is the minimum difference between the total force of the system at times i and j to ensure a converged system has been obtained.

The conjugate gradient method is a numerical algorithm for solving linear equations. The method searches for minima on the energy landscape in successive directions, thereby avoiding searching in the same direction multiple times. The search direction is obtained by moving atoms subject to the total force acting on them:

$$\overrightarrow{F}(r_i) = -\overrightarrow{\nabla}_{r_i} U^{\text{tot}}, \quad i = 1, \ldots, N; \tag{2.36}$$

where r_i is the position of atom i and U^{tot} is the total potential energy of the system. Using this method requires the calculation of the forces within the system. In the simulation of glutamate and GPE, the classical MD method of calculating forces was used to feed into the energy minimisation routines. It should be noted that, although the method for calculating the forces is derived from the classical MD approach, as the final state of the energy minimisation corresponds to a minimum in the energy landscape, the effect of temperature is not included in this method.

Parallel tempering

Realistic simulation of atomic systems using MD is limited by the time-scales available computationally. In real-life these systems evolve on time-scales of milliseconds or seconds. The longest MD simulation times are typically nanoseconds, although simple systems have been simulated for up to microseconds. Hence, discovering real processes depends greatly on the initial conformation chosen. This is a particular problem for bio-molecular systems where the energy landscape is complex and therefore transition between states is often inhibited by large potential barriers reducing the probability of transition between these states. Enhanced sampling techniques have been developed to overcome this problem.

One such sampling technique is called parallel tempering, also known as replica exchange. In this technique multiple simulations of the same system at different temperatures occur simultaneously. These parallel simulations, or 'temperers' are then periodically swapped at regular intervals throughout the simulation. This exchange facilitates the movement of high-temperature simulations, which are able to over come the potential barriers of the energy landscape, to be fed into lower temperature simulations and provide better sampling of the landscape.

Geyer was the first to develop the parallel tempering method in 1991 by using high temperature Markov Chain Monte Carlo simulations to feed into lower temperatures [97]. He described this as a 'tunnelling' between metastable states. In 1997, Hansmann applied the first parallel tempering techniques to biological systems and showed that parallel tempering with M temperers is more effective at sampling conformational space than a standard single MD simulation that runs for M times as long [98] in such instances. Sugita and Okamoto were the first to use MD parallel tempering [99]. As Monte-Carlo simulations can result in a high proportion of the attempted configurations containing over-lapping molecules, MD based parallel tempering is more useful in application to large molecule systems. Since these early examples, the effectiveness of parallel tempering has contributed to it becoming one of the most common methods for sampling the rugged energy landscape of bio-molecular systems [100, 101, 102, 103].

The algorithm for parallel tempering requires the initialisation of M replicas from the same starting configuration. Each replica or temperer performs a classical canonical ensemble simulation at a different temperature. An attempt

is made to swap each pair of temperers, i and j, between temperatures at regular intervals subject to the probability:

$$P_{ij} = min\left\{1, \quad \frac{exp(\Delta\beta_{ij}\Delta E_{ij})}{1 - exp(\Delta\beta_{ij}\Delta E_{ij})}\right\}, \tag{2.37}$$

where $\Delta\beta_{ij}$ is the difference between $\frac{1}{kT_i}$ and $\frac{1}{kT_j}$, and ΔE_{ij} is the difference in the total energy between the two replicas, E_i and E_j.

The probability of accepting a swap between two temperatures is dependent on the temperature gap between them. Eeffective use of parallel tempering requires careful choice of the initial temperature set. In addition, it is important that the temperature range includes the desired temperature for standard simulation and the top temperature is high enough to facilitate the required phase-space sampling. If the top temperature is too low then the conformation may not be able to overcome the energetic barriers associated with the system. Too high and either the temperers will be too far apart that no swaps occur or the number of temperers will be sufficient to make the problem prohibitively expensive computationally. Previous studies of bio-molecular systems have considered the range between 250 K and 700 K and provided promising results [99, 104, 105, 98]. Our work initially considers this region and attempts to identify a top temperature by looking for key differences between systems at high temperatures. This is discussed more in the set-up of the parallel tempering procedure for the glutamate system in chapter 4.

The choice of temperature set is important in order to optimise the performance of the parallel tempering simulation [103]. There is no consensus on the best choice of initial spacing, although two common choices are an equally spaced distribution [106] and geometric progressions that cluster the temperatures at the lower end of the temperature set [99, 100]. In addition there are techniques to optimise the distribution by feeding back information on the acceptance ratios of attempted swap and periodically adjusting the temperature set accordingly [107, 108, 105]. This work does not aim to devise a method applicable to all systems, but merely to utilise a reliable method for the system of interest. We have implemented a combination of the approaches above in choosing a temperature set. Having too many temperers that are closely spaced is also not ideal as this results in the conformations traversing the temperature space

too quickly. Rathore *et al.* suggest that an acceptance ratio of 20% is considered optimal [103]. In the parallel tempering work considered in chapter 4 we aimed for an acceptance probability of approximately 20%. An initial geometric distribution was initially chosen, followed by a parallel tempering simulation that included 100 attempted swaps. After this short simulation pairs of temperers that had the lowest swap acceptance ratio had their temperature spacing reduced by altering the temperatures around them, and in some cases adding an additional temperer into the set. This was repeated seven times and resulted in a temperature set of 32 temperers with acceptance ratios between 15% and 25%.

As mentioned above the spacing and range of the temperers is important to successfully executing a parallel tempering simulation. For the study of aqueous glutamate we have utilised a combination of approaches; aiming for a top temperature that enables transition between particular types of state, a temperature spacing that attempts to approach 20% acceptance ratios, and a simulation that has multiple temperers traversing the entire range of chosen temperature set. The protocol developed is discussed in more detail in chapter 4.

2.2.5 Simulation packages

Both classical and Car-Parrinello MD involve complex algorithms which have been modified and added to over many years. We have used two packages to perform the simulations required for this work. Firstly, DLPOLY_2, a numerical classical simulation tool, developed by the Daresbury laboratory is used in the simulation of glutamate systems. The second package, PINY_MD [109, 110], enables the simulation *via* CPAIMD as well as classical mechanics and has recently had parallel tempering functionality added to work alongside the classical tools.

DLPOLY_2 has the benefit of user defined empirical force-fields making it an ideal tool for the work comparing multiple water models in conjunction with the CHARMM22 forcefield. DLPOLY_2 is limited to classical potentials and therefore does not cover all aspects desired for this project. PINY_MD was developed by Martyna *et al.* in a collaboration between Indiana University, New York University and IBM Research at the T. J. Watson Institute in New York. PINY_MD has the benefit of the CHARMM22 potential being included in the package and has been adapted to support water potentials other than TIP3P. In addition PINY_MD has been set-up to work either sequentially on a

desktop computer or in parallel, as is necessary to perform parallel tempering and CPAIMD simulations. Parallel versions of DLPOLY are also available, however PINY_MD was developed in collaboration with IBM and has been optimised to work on IBM hardware such as the HPCx supercomputer used in this study.

2.3 Analysis methods

For useful information to be obtained from classical simulation the trajectories obtained need to be analysed to answer the pertinent physical questions. A combination of standard and bespoke analysis codes were exploited to achieve this. The following is a description of the various techniques used throughout the discussion, including structural and dynamical properties.

Structural properties average the instantaneous snap-shot properties of atoms and molecules across all types of that atom (where appropriate) and across all time-steps considered in a 'production' run (a simulation, starting from a fully equilibrated configuration and initial velocities, where the trajectory is sampled purely for analysis of the system). Dynamical properties of the system consider the changes in the configuration through time. Dynamics give a measure of the freedom of movement of the system and are therefore important in characterising the effects of solutes on a solution.

2.3.1 Structure factor

The static structure factor is mainly used in this instance as a tool for comparison with experimental data. Experimentally the structure factor describes the elastically scattered radiation and can be obtained directly from X-ray, electron and neutron diffraction data, after correcting for background and recoil effects [111]. Comparison of the structure factor, $F(Q)$ enables a clear measure of the quality of the simulated solution structure compared to experiment.

A common approach is to compare total distribution functions obtained from experiment and reverse Monte Carlo techniques such as EPSR (Empirical Potential Structure Refinement) [112, 113]. EPSR is a modelling technique that is constrained by experimental data, allowing direct comparison with experiment, however it is still a model and therefore comparing simulation results to EPSR removes the benefit of comparing directly with experiment.

The traditional approach of obtaining a structure factor from computational simulation involves a Fourier transform of the number density of atoms in the system. Neutron diffraction measures the structure factor, $S(Q)$ which is Fourier transformed into $G(r)$, the total pair correlation function that is a linear combination of the individual correlation functions:

$$G(r) = \sum_{ij} b_i b_j c_i c_j g_{ij}(r), \tag{2.38}$$

where b_i is the coherent neutron scattering length and c_i is the mole fraction of species i. $g_{ij}(r)$ is the site-site pair distribution, which can be obtained from experiment by isotopic substitution. Commonly in Molecular Dynamics the static structure can be obtained by the continuous sine transform of

$$H(r) = \sum_{ij} b_i b_j c_i c_j [g_{ij}(r) - 1] \tag{2.39}$$

This is only valid in the thermodynamic limit where the total number of particles in the system, N tends to infinity. At low Q for a finite system, such as those studied by computational simulation, the necessary truncation of the transform means that error is inherent in the calculation.

An alternative approach has been described by Troitzsch *et al.* [106] where direct comparison with the experimentally obtained structure factor is possible, after correcting for nuclear recoil, background noise and multiple scatterings (see ref [111]) by directly calculating the structure factor from simulation. Determination of the structure factor in this way provides a direct, unambiguous measure of the quality of the solution structure obtained from the simulation as it enables direct comparison with data from neutron diffraction experiments, as discussed in refs [106] and [114].

The full derivation of this method is provided in refs [106] and [114], and as such we will not reproduce it in its entirety here. The following is a concise presentation of the derivation in the supplementary material by Tulip *et al.* [114].

In neutron diffraction the total scattering cross-section for an N-particle

system is defined as:

$$S_{\text{neutron}}(Q) = \frac{1}{N\Omega_{\hat{Q}}} \left\langle \sum_{\hat{\mathbf{Q}}\prime}^{\prime} |n_{\text{neutron}}(\mathbf{Q}\prime)|^2 \right\rangle, \tag{2.40}$$

where $\Omega_{\hat{Q}}$ is the number of reciprocal space lattice vectors, $\hat{\mathbf{Q}}$, the single particle atomic density function is;

$$n_{\text{neutron}}(\mathbf{Q}) = \sum_j b_j e^{i\mathbf{Q}.\mathbf{r}_j}, \tag{2.41}$$

which is weighted by the *coherent neutron scattering length* of each particle b_j. This length can decomposed into partial scattering lengths;

$$S_{\text{neutron}}(Q) = \sum_{\alpha\beta} c_\alpha c_\beta b_\alpha b_\beta S_{\alpha\beta}(Q). \tag{2.42}$$

Obtaining the partial cross sections, or structure factors, from simulation enables direct computation of the total structure factor. We consider the decomposition of the related total distribution function into partial distribution functions:

$$G(r) = \sum_{\alpha\beta} c_\alpha c_\beta b_\alpha b_\beta g_{\alpha\beta}(r). \tag{2.43}$$

The partial radial distributions, $g_{\alpha\beta}(r)$, for a multicomponent system (note that the full derivation from a single component system is given by Tulip *et al.* [114]) are;

$$g_{\alpha\beta}(r) = \left[\frac{1}{4\pi\rho r^2 c_\alpha c_\beta}\right] \left[\left\langle \sum_{i_\alpha=1}^{N_\alpha} \sum_{j_\beta=1}^{N_\beta} \sum_{l\neq 0; i_\alpha = j_\beta} \delta(|\mathbf{r}_{i_\alpha} - \mathbf{r}_{j_\beta} - \mathbf{hl}| - r) \right\rangle \right], \tag{2.44}$$

where the mole fraction for species α is $c_\alpha = \frac{N_\alpha}{N}$, and N is sum of all N_α. The density is $\rho = \frac{N}{V}$ and $\mathbf{r}_{i_\alpha}$ is the position of the i^{th} atom of type α. $\mathbf{h}$ is a matrix that describes a parallelpiped of volume V, which in turn is described by the lattice indices $\mathbf{l}$ such that $\mathbf{h}$ = boxlength $\times$ $\mathbf{l}$. In addition the matrix $\mathbf{h}$ provides the magnitude of the reciprocal space lattice vectors, $bf\hat{Q}$, as $|2\pi\hat{\mathbf{Q}}\mathbf{h}^{-1}|$. As expected for a radial distribution function this approaches unity as $r \to \infty$ as radial distribution functions are normalised by the ideal gas density of the

particles under consideration. From this the partial structure factor, $S_{\alpha\beta}$ can be defined as:

$$c_\alpha c_\beta S_{\alpha\beta}(r) \equiv \frac{1}{N\Omega_{\hat{Q}}} \left\langle \sum_{\hat{\mathbf{Q}}\prime}^{\prime} n_\alpha(\mathbf{Q}\prime) n_\beta^*(\mathbf{Q}\prime) \right\rangle, \tag{2.45}$$

where

$$n_\alpha(\mathbf{Q}) = \sum_{j_\alpha=1}^{N_\alpha} e^{i\mathbf{Q}.\mathbf{r}_{j_\alpha}} \tag{2.46}$$

In a finite system this can be written as;

$$g_{\alpha\beta}(r) + \left[\frac{\delta(r)}{2\pi r^2 \rho c_\alpha c_\beta}\right] \delta_{\alpha\beta} c_\alpha = \frac{1}{N} \sum_{\hat{Q}} \Omega_{\hat{Q}} S_{\alpha\beta}(Q) \frac{\sin(Qr)}{Qr}, \tag{2.47}$$

where $\delta_{\alpha\beta}$ is the Kronecker-delta function:

$$\delta_{\alpha\beta} = \begin{cases} 1, & \text{if } \alpha = \beta \\ 0, & \text{if } \alpha \neq \beta \end{cases} \tag{2.48}$$

This allows the direct calculation of the static structure factor from the partial radial distribution functions that can be obtained from simulations. The partial distribution functions, also referred to as radial distribution functions, are also inherently useful in characterising system structure. These are considered in more detail in the following section.

2.3.2 Radial distribution functions

Radial distribution functions, or pair distribution functions are commonly used to characterise the local structure of a liquid. These are particularly useful because neutron and X-ray scattering can provide the total distribution function of simple fluids *via* Fourier transform of the structure factor. In addition $g(r)$ is directly related to theories of state, thereby enabling calculated numerical values of $g(r)$ to be compared to theoretical predictions.

The function $g_(\mathbf{r}_i, \mathbf{r}_j)$ (or $g_(r_{ij})$, $g(r)$) is related to the probability of finding a particular pair of atoms a distance r apart:

$$g(r) = \rho^{-2}\langle\sum_i\sum_{j\neq i}\delta(\mathbf{r}_i)\delta(\mathbf{r}_j-\mathbf{r})\rangle \quad (2.49)$$

$$= \frac{V}{N^2}\langle\sum_i\sum_{j\neq i}\delta(\mathbf{r}-\mathbf{r}_{ij})\rangle \quad (2.50)$$

where ρ is the density of atoms in the system, V is the volume and N is the total number of atoms.

Calculation from simulation only requires the ratio of the number density of a particular species of atom ρ_i at a distance r from an atom j and the density at a distance r from an atom in an ideal gas of the same overall density averaged across the entire length of the simulation.

For a system such as glutamate or GPE in aqueous solution it is prudent to consider the individual atomic species in order to gain the greatest understanding of the system. In this instance numerical routines consider pairs of particular species in the calculation of $g(r)$, rather than all atoms within the system. For example g_{O-H} is the radial distribution function between all atoms of type O and all atoms of type H within the system under consideration. In the analysis used for this work different atoms are given different labels, in particular distinctions are made between multiple atoms of the same species depending on where in a molecule they reside.

As the distribution function is a measure of the probability of finding an atom of type α at a particular distance from an atom of type β, the co-ordination number ($n_{\alpha\beta}$, the expected number of atoms of type α around an atom of type β) can also be extracted from this analysis. This can be computed from the integral of the number density in the range from 0 to r_1;

$$n_{\alpha\beta} = 4\pi\int_0^{r_1} r^2 g_{\alpha\beta}(r)\rho dr, \quad (2.51)$$

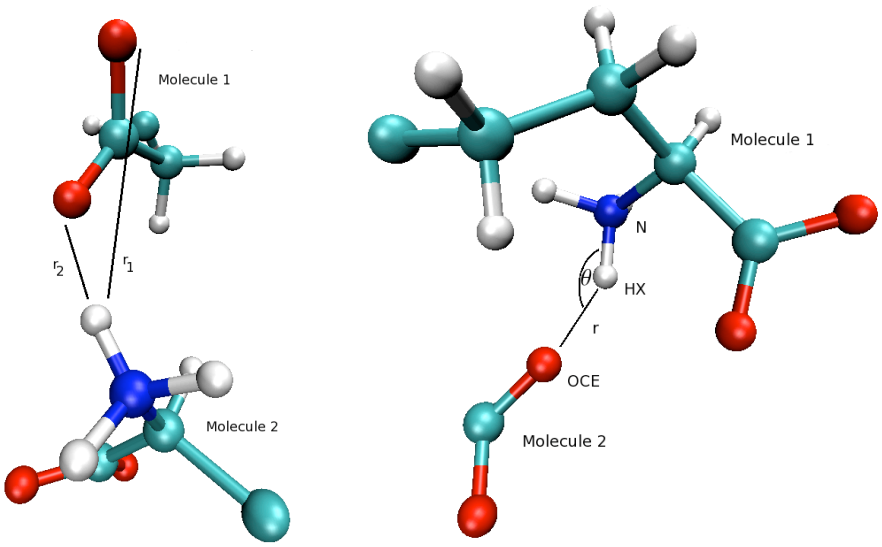
where the average density of α atoms around β atoms is ρ.

Computationally, the radial distribution function between atoms α and β is calculated by summing up the radial distance between all atom pairs at all time steps and producing a histogram for values of r with bin-width dr. This is then normalised by dividing by the ideal gas density of for the N atoms in the system

of volume V, N/V, and the total number of time-steps used. It should be noted that $g_{\alpha\beta}(r) = g_{\beta\alpha}(r)$, but that $n_\beta^\alpha \neq n_\alpha^\beta$.

2.3.3 Joint radial-radial and radial-angular distribution functions

Two dimensional, or joint distribution functions, consider the radial distribution between two atoms, i and j, coupled with a second variable linked to one of the atoms, such as the angle between them with respect to another atom. In this work the radial-radial, $g(r_1, r_2)$, and the radial-angular, $g(r, \theta)$, correlations are considered [33]. A diagram illustrating the type of contacts plotted in the 2D distributions is shown in figure 2.3.



(a) Radial-Radial (b) Angular-Radial

Figure 2.3: Diagrams of the 2.3(a) radial-radial and 2.3(b) radial-angular distribution functions.

For the radial-radial distribution (figure 2.3(a)), the plot correlates the radial distances r_1 and r_2 for atom α, β and γ, where r_1 is the distance between atom α in molecule 1 and atom β in molecule 2, and r_2 is the distance between atom α in molecule 1 and atom γ in molecule 2. The angular-radial distribution for atoms $\theta_{\alpha-\beta-\gamma}$, correlates the radial distance $\alpha - \beta$ with the angle $\alpha\hat{\beta}\gamma$, as shown in figure 2.3(b).

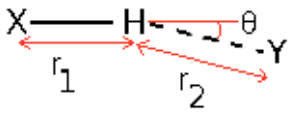


Figure 2.4: Schematic of a hydrogen bond.

These quantities are calculated computationally as the ratio of the number density of atoms at a given r and r or θ, with the ideal gas density. From a computational viewpoint, this type of calculation is much more intensive as it greatly increases the number of histogram bins required.

Integrating across the second variable (either r_2 for the radial-radial distribution functions or θ for the angular distribution functions) reproduces the radial distribution function.

2.3.4 Hydrogen-bonded cluster distribution

Glutamate, GPE and water molecules can all form inter-molecular bonds. When bonds are formed between two molecules of the same species they can be considered a 'cluster' of molecules. Statistical analysis of the clusters of glutamate and GPE in their respective aqueous solutions provides a further characterisation of the liquids.

To define a cluster, a description of a hydrogen bond is first required. A hydrogen bond is defined by two bonds, a donor X-H bond, that is a short, intra-molecular interaction, and an acceptor, H$\cdots$Y, a longer inter-molecular interaction, as shown in figure 2.4. X and Y are electronegative atoms that bond to a proton (H atom). According to Brown *et al.* the 'strong' X-H covalent bonds (r_1 in figure 2.4) are defined by the very short radial interaction (1.4-1.7 Å), while the weaker H-Y hydrogen bonds (r_2 in figure 2.4) have a longer radial interaction distance (1.7-2.1 Å) [115]. According to Pedersen *et al.* the hydrogen-bond angle for water molecules (θ in figure 2.4) varies from 0° for strong H-bonds, to 30° for the weakest H-bonds [116]. These are only characteristics of H-bonds, they do not define them. This qualitative picture of hydrogen bonds can be determined from MD and Monte Carlo simulations, however, there is no unequivocal definition of this interaction [117], particularly as the interaction is a continuous function of energy that depends on the surrounding thermodynamic conditions.

To enable analysis of hydrogen-bond behaviour, two sets of criteria have been used historically to facilitate the analysis of computer simulations of hydrogen-bonded fluids, energetic and geometric. It has been shown that energetic and geometric criteria lead to similar results [118, 119, 120, 121]. Geometric criteria are the simplest definition to implement in MD simulations, In this work we have chosen geometric criteria, such as those used in dynamical analysis of hydrogen bonds by Padrò *et al.* [122].

A cluster is defined as a set of molecules of the same species forming hydrogen-bonds between molecules, so that they are all connected *via* by at least one bond to the group, at the same time (i.e. the same trajectory snapshot from any given simulation). Two molecules are considered hydrogen-bonded if they satisfy *all* of the following criteria:

1. $r_{XY} < r_{XY-cut}$ - the radial separation of atoms X and Y is less than a minimum separation r_{XY}.

2. $r_{HY} < r_{HY-cut}$- the radial separation of atoms H and Y is less than a minimum separation r_{HY}.

3. $\theta_{X-H-Y} < \theta_{cut}$ - the angle of the bond with respect to the linear configuration is less than a cut-off θ_{cut}.

An angular cut-off is included as Guárdia *et al.* noted that hydrogen-bonds normally approach a linear configuration [123], and the absence of near-linear configurations indicated 'non-physical' bonds that were energetically unfavourable and short-lived indicating that the association was transitory rather than the formation of a hydrogen bond. For the cluster histograms we have used an angular cut-off of 30°, as this has been used by other groups in the examination of complex liquids. We have also plotted the angle distribution of hydrogen bonds determined by just the radial separations, and found that in the glutamate solutions the angles are almost always $\theta \leq 10°$. This is discussed in more detail in chapter 4.

A cluster-frequency distribution graph is then obtained by analysing the proportion of clusters of a particular size, as a percentage of the total number of clusters. To identify the number of clusters of a certain size for a single time-step the following algorithm is used:

1. Loop round all molecules of the species under consideration. For each molecule α, fill an array ('bond-list') with a list of the indices of all molecules that molecule α is bonded to. Repeat until a list for all molecules has been completed.

2. Create a 'cluster' array which lists the constituents molecules for each cluster. Fill by looping round the 'bond-list' array. For each molecule, α, find if index α has already been entered into the cluster array, if true then add the molecules listed in bond-listα to the correct 'cluster' entry. If false then create a new entry in 'cluster' filling it with index α and all the molecule indices listed under bond-list(α). Repeat for all entries in 'bond-list'.

3. Create array 'size-count', equal in length to 'cluster'. Fill so that size-count(i)=number of molecules listed in cluster(i)'.

4. Loop round 'size-count' incrementing a histogram of cluster sizes for each entry of size-count. For each example, if size-count(j)=4 then history(4) would be incremented by 1.

5. Repeat steps 1-4 for all time-steps.

An alternative approach, not considered here, would be to calculate the distribution the number of molecules contained in a particular cluster size as a fraction of the total number of molecules available, resulting in a very different distribution.

2.3.5 Ramachandran plots

Ramachandran maps are a method used to visualise the dihedral angles of peptide residues [124]. The Ramachandran maps were developed by Gopalasamudram Narayana Ramachandran and Viswanathan Sasisekharan in 1963 to provide a concise method for understanding the various conformations of polypeptide chains. A Ramachandran map shows the distribution of the two degrees of freedom of a peptide bond (the backbone of a polypeptide). These degrees of freedom are the dihedral angles ψ and ϕ that describe the orientation of any peptide bond (see figure 2.5).

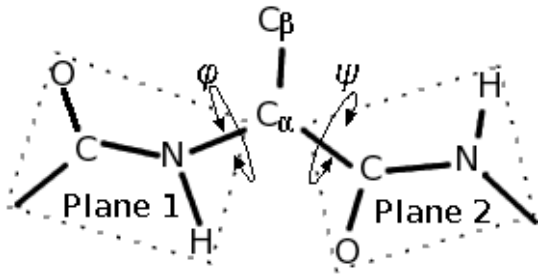
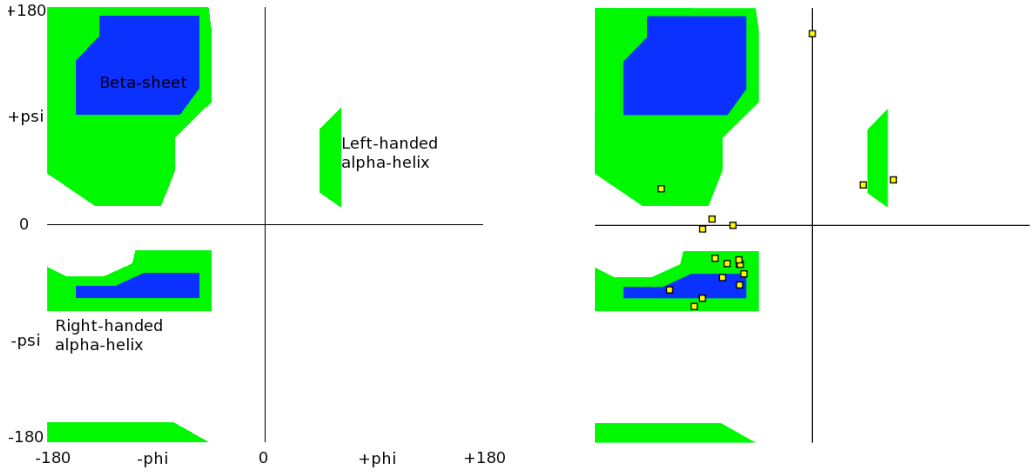


Figure 2.5: A peptide bond showing the two degrees of freedom, the dihedral angles ϕ and ψ.



(a) Diagram of a Ramachandran Map (b) Example Ramachandran Map

Figure 2.6: 2.6(a) Diagram showing the allowed regions of a Ramachandran map. 2.6(b) Example of a Ramachandran map showing the distribution of torsional angles of the 55-72 fragment of staphylococcal nuclease obtained from the protein databank [2].

Figure 2.5 shows that the O-C-N-H on one side of a peptide bond lies in a two-dimensional plane. The dihedral (or torsional) angles measure the relative orientation of the plane with respect to the C_α atom. The two-dimensional plane of the Ramachandran map is a projection of the 3-dimensional toroidal space that the dihedral angles occupy. This results in a distorted and discontinuous plot of a continuous domain visualising the function f:$[-\pi, \pi) \times [-\pi, \pi) \rightarrow \mathbb{R}_+$. An example of this type of map is shown in figure 2.6(a).

Ramachandran initially analysed possible steric clashes within the protein backbone, thereby eliminating some regions of the map. Figure 2.6(a) shows blue and green coloured regions that correspond to the sterically allowed regions.

The white regions correspond to sterically disallowed configurations. These were identified by treating the atoms as hard spheres with a radius consistent with the corresponding Van der Waal radii. When the angles ψ and ϕ caused the spheres to collide the configuration was deemed unphysical. There are three main regions, known as α, α_L and β, where the values of $\phi - \psi$ were allowed. These regions correspond to the α-helix, the α-left helix and the β-sheet backbone secondary structures. The regions of sterically allowed dihedral angles have changed over successive years, but on the diagram shown are the regions of sterically allowed under a weak constraint (green) and under stricter constraints (blue), where a larger *van der Waal's* radius is used to represent the atoms. The green regions are the only way to enable a left-handed helix (in the top right quadrant of the map).

An example Ramachandran map produced by the molecular analysis program VMD [125] is shown in figure 2.6(b). This plots the torsional (dihedral) angle correlations on the backbone of the 55-72 fragment of staphylococcal nuclease [5] (the structure has been obtained from a PDB file available from the protein databank [2]). Points that lie on the axis of the plot indicate the N and C-terminal residues for each subunit, or residue, of the protein fragment.

2.3.6 Hydrogen bond dynamics

Defining a hydrogen bond, as discussed in section 2.3.4, enables characterisation of the structure of complex liquids. Understanding the types of interaction preferred, requires the consideration of the longevity of hydrogen-bonds as well as the instantaneous distribution.

As with the hydrogen-bond cluster distribution we have chosen to use geometric criteria to define a hydrogen bond, including the two radial cut-offs and an angular cut-off as described in section 2.3.4. This is applied to all bonds, solute-solute, solute-solvent and solvent-solvent. In addition to the complex issue of how to define a hydrogen bond, considering the dynamics of a hydrogen bond raises additional questions. Librations in the bonds (ultra-fast movements where the bond continually flexes or vibrates back and forth around a central point) can result in temporary breaking of the bond (according to the stated criteria), which are subsequently rapidly re-formed such that the bond criteria are fulfilled again. As emphasised in section 2.3.4 the definition of a hydrogen bond is somewhat

arbitrary. If a bond temporarily breaks subject to these criteria that does not necessarily mean that the bond has truly broken in a physical sense. Indeed, a sufficiently rapid 're-formation' would suggest that the initial criteria are too stringent. Determining how to resolve this problem is not straightforward.

Rapaport introduced a method, now standard, to determine the relaxation times from history-dependent hydrogen-bond auto-correlation functions [61]. Since then two types of auto-correlation have been consistently used to characterise lifetimes of hydrogen-bonds. The first method considers history-dependent correlations functions, $c(t)$, where the hydrogen bond must be continuously bonded from time 0 to time t for the function to be none-zero at t. The second method considers a history-independent auto-correlation function, where $c(t)$ is non-zero at t if the bond is formed at time 0 and at time t but has no dependence on what occurs in between these two times. History-dependent auto-correlation functions depend on the definition of the hydrogen bond, while history-independent are not necessarily the best analysis of bond relaxation times due to their history-independent nature [126, 127].

An alternative method is to use history-dependent auto-correlation functions but to allow temporary breakages of the bonding criteria, provided that the bond is 're-formed' within a specified time-span [128]. This is the method used in this work as it enables the use of a history-dependent correlation, as well as flexible criteria for the hydrogen-bond.

History-dependent auto-correlation functions

To enable computation of a correlation function for a hydrogen bond, $c_{HB}(t, t^*)$, a binary operator is defined:

$$\eta_{ij}(t) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are continuously bonded between 0 and } t, \\ 0 & \text{otherwise.} \end{cases} \tag{2.52}$$

where 'bonded' refers to the hydrogen-bond criteria discussed in section 2.3.4. Using this binary operator, an auto-correlation function is then constructed:

$$c_{HB}(t) = \frac{\langle \eta_{ij}(t).\eta_{ij}(0) \rangle}{\langle \eta_{ij}(0)^2 \rangle} \tag{2.53}$$

The function $c_{HB}(t)$, therefore represents the probability that a pair of atoms, forming a hydrogen-bond at time 0 are still bonded at time t. This is then extended to take into account the ability to temporarily break hydrogen-bonds that subsequently re-form within a specified time, t^*.

$$c_{HB}(t, t^*) = \frac{\langle \eta_{ij}(t, t^*).\eta_{ij}(0) \rangle}{\langle \eta_{ij}(0)^2 \rangle} \tag{2.54}$$

In this instance $\eta_{ij}(t, t^*) = 1$ if the hydrogen bond criteria between atoms i and j have been met between times 0 and t and has been continuously broken for no longer than t^* during this period.

Starr *et al.* showed that a relaxation time can be obtained by fitting $c(t)$ to an exponential function [129]. Nieto-Draghi *et al.* showed that $c(t, t^*)$ produced an asymptotic exponential decay function and we are therefore able to extract a relaxation time, or hydrogen-bond lifetime τ_{HB} from this function [128]:

$$c_{HB}(t, t^*) \sim exp(-\frac{t}{\tau_{HB}(t^*)}) \tag{2.55}$$

As with the definition of the auto-correlation and the hydrogen-bond criteria, there is no clear answer as to whether this method satisfies the physical complexity of obtaining a hydrogen-bond lifetime. Other methods would consider how long it takes for the correlation function to halve as the relaxation time. As this part of the work is primarily interested in changes between the systems under consideration in this study alone, consistency within this work is the main aim, so this method is deemed sufficient.

Computationally, calculation of the auto-correlation function requires tracking each pair of atoms that have the potential to form a hydrogen-bond through each time-step, and filling a three-dimensional tensor subject to whether the criteria are met for atoms i and j at time t. A histogram can then be constructed determining the number of pairs that had a lifetime of t in length. This is then normalised and fitted to $exp\left(-\frac{t}{\tau_{HB}(t^*)}\right)$.

Diffusion

Diffusion is a measure of the rate of transport through space due to the thermal processes within a medium and in the absence of external influences inducing flow.

Fick's Law of diffusion links the diffusive flux, j to the concentration gradient, c;

$$\mathbf{j} = D\nabla c \tag{2.56}$$

where D is the diffusion coefficient.

In the solutions considered in this study, there are many molecules of the same species. We therefore consider self-diffusion, where a single labelled particle travels through a solution of otherwise identical particles. In this instance the Einstein diffusion relation combines Fick's Law with conservation of mass to enable the self-diffusion coefficient, D, to be equated to the mean square displacement of the particle:

$$D = \lim_{t\to\infty} \frac{1}{6t} \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle \tag{2.57}$$

where the $6t$ takes into account all three dimensions.

Over very small time-scales molecular movements are quadratic in nature as there are no collisions. However, after a longer time collisions occur with other particles and thereby change the mean squared displacement of the molecule to a diffusive pattern. The diffusion coefficient requires the long-time linear portion of the mean squared displacement, and hence the equation requires the limit $\lim t \to \infty$ is approached as closely as practicable.

The mean squared displacement $\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle$ can be easily extracted from MD simulation trajectory files by observing the progress of individual particles, or molecules, over the length of the simulation. D is calculated from the full simulation time mean squared displacement.

Chapter 3

The Structure of Aqueous Glutamic Acid

3.1 Introduction

This chapter discusses the structure of glutamate molecules (hereafter referred to as 'glu') in an aqueous solution with sodium ions. We consider both classical simulations (section 3.3) and a 'first principles' approach (section 3.4). The classical simulations focus on the glu-glu interactions as this has not been studied before to date.

The classical simulations parameterise the glu molecules and sodium ions with the CHARMM22 classical forcefield [71]. This is considered in conjunction with three different water potentials, F3C [81], SPC/E [82] and TIP3P [72]. The 'first-principles' calculation of the glu system uses Car-Parrinello *ab initio* Molecular Dynamics (CPAIMD, also referred to as CPMD as it is not strictly speaking an *ab initio* method) [88].

In order to aid the following discussion the atoms in the glu molecule are labelled as per the scheme shown in figure 3.1 and not by their chemical name. This enables the distinction between the two types of carboxyl groups on the glu-molecule. The side-chain carboxyl oxygens are labelled O_{CE} and the C_α oxygens, O_C. In addition to the glu-molecule atoms, the water molecule is constructed from two H_W and one O_W atom, and the sodium ion is referred to as Na^+.

A selection of the results from the classical simulations have been published in the Physical Chemistry Chemical Physics (PCCP) journal by Collis *et al.* in

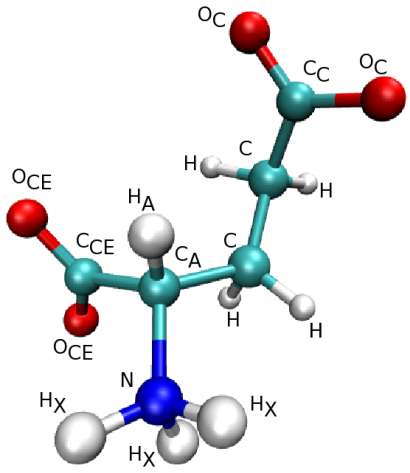


Figure 3.1: The labelling convention of the glu molecule. C, C_A, C_C and C_{CE} are all carbon atoms; C_{CE} is the side-chain carboxyl-carbon and C_C the C_α carboxyl-carbon. H, H_A and H_X are all hydrogen atoms, and O_{CE} and O_C are oxygen atoms.

2010 [4].

3.2 Methods

Four systems of aqueous glutamate in solution with water molecules and sodium ions are presented in the following chapter. All four systems are studied at a glu:sodium:water concentration of 1:1:29, enabling direct comparison with experimental neutron diffraction data of a 1:1:29 solution obtained by McLain *et al.* in 2006 [3].

The first three systems modelled use atomistic classical MD and are compared in section 3.3. The CHARMM22 classical force-field is used to parameterise the glu molecules and the sodium ions, in conjunction with one of three water potentials, the rigid TIP3P [72] and SPC/E [82] models, and the flexible F3C [81] model. These three systems all contain 20 glutamate molecules, 20 sodium ions and 580 water molecules.

The initial configuration for the simulations was constructed by placing a conformation of the glu-molecule, obtained from the protein databank, on a lattice within a box 100 Å in length, twenty times. Each system was then solvated with 580 water molecules, creating two boxes with water molecules, the first in

the TIP3P geometry and the second in the SPC/E geometry. The second box was then replicated, because the SPC/E geometry of water is the equilibrium geometry for the flexible model, F3C.

The fourth system, discussed in section 3.4, and set up as the systems described above, uses CPAIMD to perform a 'first principles' calculation on a system of two glu molecules, two sodium ions and 58 water molecules in a cubic box of length 30 Å, which is then subjected to an NPT simulation to obtain the correct density.

The equilibration process of the three classical systems involved subjecting each system to a 25 ps NPT classical MD simulation using the Berendsen thermostat/barostat [78], using 0.5 fs time-steps, atmospheric pressure and at 300 K, using DLPOLY_2. This resulted in the volume of the boxes decreasing until the correct theoretical density of the system had been achieved, at a cubic box length of 27.3 Å. Each system was then subjected to a further classical 1 ns NVT equilibration, using the Nosé-Hoover thermostat [77] at constant temperature and constant volume. The structural data was collected from a 1 ns NVT production run, using 0.5 fs time-steps at 300 K, with the configuration being sampled every 50 fs.

The CPAIMD system was also initially subjected to a *classical* 100 ps NPT equilibration, using the Berendsen thermostat/barostat, until the target density was achieved, resulting in a cubic box of 12.8 Å in length. This was followed by a 1 ns classical NVT (Nosé-Hoover) simulation, using a 1.0 fs time-step at 300 K using the PINY_MD simulation package, where 1 fs time-steps are possible due to the use of a Nosé-Hoover chain of length 2. For both the NPT and NVT classical equilibration procedures the CHARMM22 and SPC/E force-fields were used the characterise the atoms in the system. The resulting configuration was then subjected to CPAIMD simulations at 300 K using the HPCx UK National Supercomputing Facility [130]. The CPAIMD simulations use norm conserving pseudopotentials, the gradient-corrected Becke-Lee-Yang-Parr [94] (B-LYP) approximate density functional and the B-LYP plane wave basis set as this has been shown to effectively reproduce pure water [95, 96]. The plane-wave expansions of the wave-functions were restricted to reciprocal lattice vectors with a magnitude less than 70 Ryd, and the Fourier components of the charge density were restricted by a cut-off of 280 Ryd. This was all implemented using a Nosé-

Hoover chain thermostat [79] of length 2. Initially a 3.6 ps CPAIMD calculation was used to further equilibrate the system, removing the constraints that may have been imposed by the classical simulation. The structural data studied in this chapter was then obtained from a further 20 ps production run, using 0.15 fs time-steps, sampled ever 10 steps, resulting in 33,333 configurations.

To consider the effects of the box sizes used in the standard classical and CPAIMD simulations, two further simulations were carried out. In both instances the simulations used a solution at the same 1:1:29 concentration. The first system, used to assess the finite size effects experienced by the classical simulations, contained 160 glu molecules, 160 sodium ions and 4640 water molecules, with sides of 54 Å in length. The second system, used to assess the influence of the small box size used in the CPAIMD simulations, used the input configuration obtained for the CPAIMD simulations. This was subjected to the same equilibration and production run procedures as the three standard-sized classical simulations, including the full-length 1 ns production. Due to the size, the large box was only subjected to a 0.5 ns NVT equilibration and 250 ps production run. Both finite-size check simulations were carried out using DLPOLY_2 to enable direct comparison with the standard sized simulations.

In our analysis of the classical production runs two further simulations were performed to aid the discussion (this is discussed in more detail in the following analysis). The first simulation considered a single glu-molecule and sodium ion in gas phase in order to establish the dependence of the structure factor results on the presence of water molecules. The second changed the initial configuration of the glu-molecules in the standard classical simulation, placing them closer together in an aggregate before surrounding the molecules with water. This was achieved by placing the 20 glu molecules on a lattice, with the sodium ions, then subjecting this to a 200 ps NPT simulation at atmospheric pressure and 300 K. The box size was then increased to 50 Å and the glu and sodium ions were surrounded by water molecules, before being subjected to the same equilibration procedures as the main systems, and using the F3C potential for the water molecules. This ‘clustered-run’ enabled analysis of the dependence of initial structure of the system on cluster size throughout the classical simulations considered. For both these systems the aim was to provide useful comparisons of the structure factor for the gas-phase simulation and of the cluster-size histograms

for the clustered-run, and therefore all other aspects of the simulations were kept the same as the other classical systems, using the CHARMM22 potential, F3C water molecules in the clustered run, and the same run-time parameters, including a full 1 ns production run.

In the following discussion, the following labels are used to refer to the systems:

- EXP: Experimental data obtained by McLain *et al.* [3] for a 1:1:29 solution of glu, sodium ions and water.
- Large Box: 250 ps simulation of 160 glu molecules with sodium ions and F3C parameterised water molecules.
- F3C: Production run of CHARMM22 glu and sodium ions with F3C parameterised water molecules.
- SPCE: Production run of CHARMM22 glu and sodium ions with SPC/E parameterised water molecules.
- TIP3P: Production run of CHARMM22 glu and sodium ions with TIP3P parameterised water molecules.
- CPMD: Car Parrinello simulation of 2 glu, 2 sodium ions and 58 water molecules.
- Small Box: Small classical box of 2 glu, 2 sodium ions and 58 SPC/E parameterised water molecules.
- GAS: Single glu molecule and sodium ion simulated under the same conditions as the classical production run systems.
- Clustered: Classical simulation of CHARMM22 glu/sodium ions and F3C water using an artificially aggregated cluster of glu-molecules for the initial configuration.

3.3 Structural analysis of classical simulations of aqueous glutamate

3.3.1 Structure Factor determination and validation of simulation

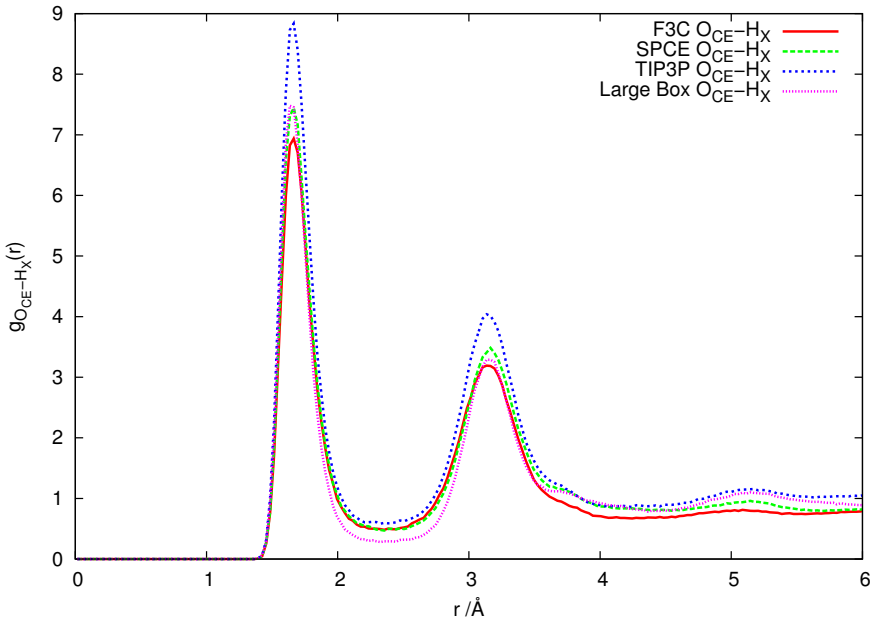
A large system, eight times the volume of the standard sized simulations, was considered to assess any finite-size effects the classical simulations may experience. Figure 3.2 shows the RDFs for the two types of glu-glu interactions between the amine hydrogen (H_X) and the two species of carboxyl oxygen, O_{CE} and O_C. Given the relatively small number of glu molecules in the system, coupled with their large size, any finite size effects should manifest in these interactions first.

Comparison of the RDFs between the large box and the three smaller boxes of the glu solution shows no significant deviations between the large box and the three production runs (see figure 3.2). Further comparison is provided by the co-ordination numbers for the two glu-glu interactions (shown in table 3.1. This shows only a minimal increase in the co-ordination numbers for the large box system compared to the standard simulation. It can assumed that the smaller size of the production run systems is not affecting the physical nature of the simulations.

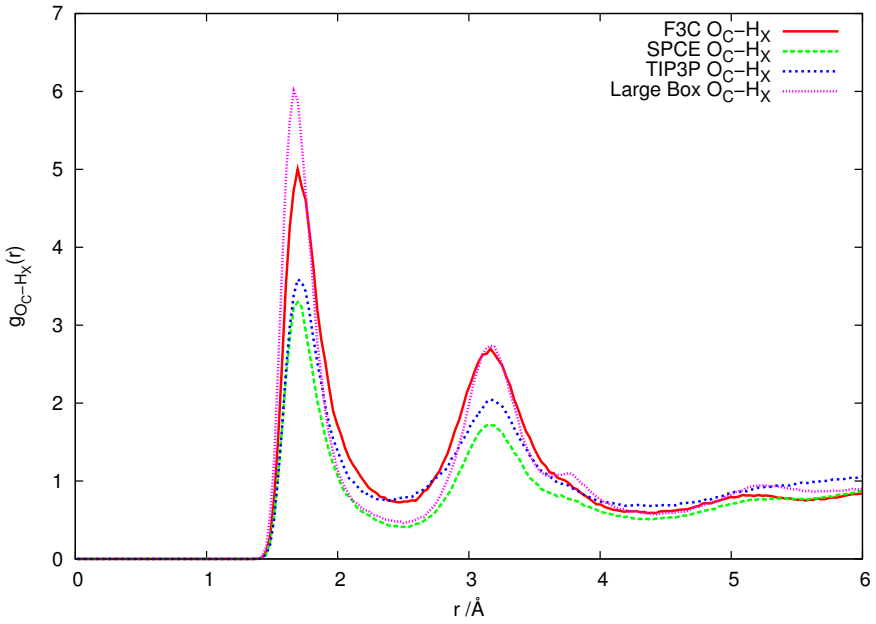
RDF	F3C standard		F3C Large	
	r_{min} (Å)	$n_\alpha^\beta(r)$	r_{min} (Å)	$n_\alpha^\beta(r)$
$g_{H_X-O_{CE}}(r)$	2.43	0.32	2.43	0.44
$g_{H_X-O_C}(r)$	2.08	0.21	2.08	0.28

Table 3.1: Co-ordination numbers n_α^β at a radial distance r_{min} for the two species of glu-glu bonds in the large box and a standard F3C system.

The second check on the accuracy of the classical MD simulations considers the comparison of the static structure factor, obtained directly from the classical simulations, with the experimentally obtained structure factor from the work by McLain *et al.* (see figure 3.3). The method of calculating the structure factor, $F(Q)$ from a simulation, as described by Troitzsch *et al.* [106], enables direct comparison with experiment as $F(Q)$ can be directly obtained from neutron diffraction data (after correcting for nuclear recoil, background noise



(a) $g_{O_{CE}-H_X}(r)$)



(b) $g_{O_C-H_X}(r)$

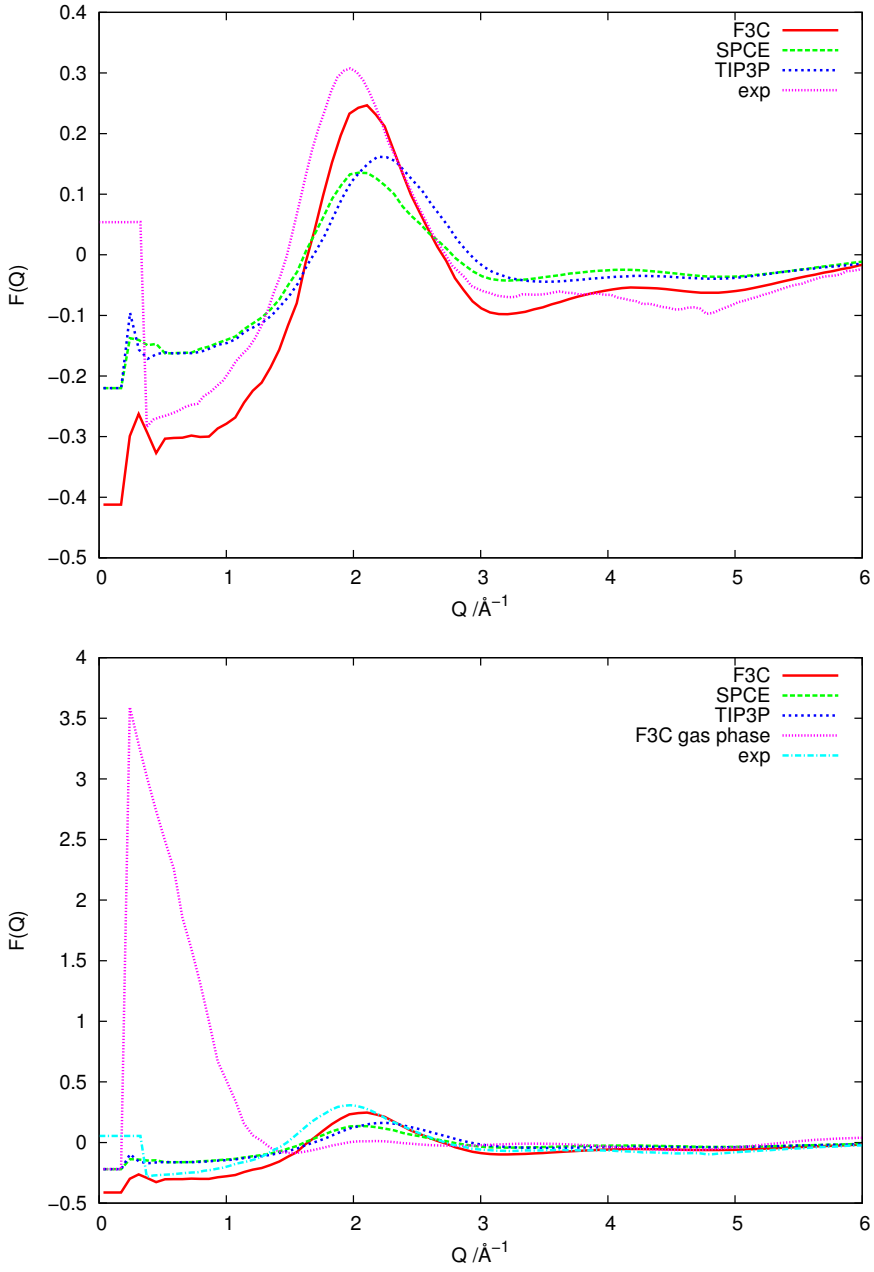
Figure 3.2: Comparison of the glu-glu RDFs in the large-box simulation and the three standard-sized simulations.

and multiple scatterings as described in ref [111]). A more common approach would be to compare our simulation to EPSR-derived distribution functions, however, EPSR is a modelling technique constrained by experimental results, and so does not provide the direct comparison to experiments that calculting $F(Q)$ directly enables. Figure 3.3 shows the experimentally obtained structure factor by McLain *et al.* [3] for a glu solution at the same concentration as the classical simulations under discussion here. In neutron scattering intensities of H-nuclei has a large inelastic component. This results in a large background noise component that is difficult to correct for. However, using d-glutamate in D2O alleviates this problem as comparison of the data sets enables contrast variation. Therefore we show $F(Q)$ calculated from the simulation using a scattering length of $b_D = 6.671fm$ rather than the hydrogen scattering length of $b_H = -3.741fm$, so that we can compare with the experimental data by McLain *et al.* for d-glutamate in D2O . Figure 3.3(a) shows $F(Q)$ from neutron diffraction and the three classical solutions and figure 3.3(b) shows these four systems and the F3C gas-phase simulation data (the reader should not the change of scale on the second plot).

All three classical simulations reproduce the key features of the experimental structure factor, and show reasonable agreement with the experimental structure. This indicates good reproduction of the intra-molecular structure in the simulations. The F3C system provides the closest fit to the neutron diffraction data, although all four simulated systems (including the gas-phase) underestimate the peak at $Q \sim 2\text{Å}^{-1}$. However, the F3C is the only system to under-estimate the trough at $Q \sim 3\text{Å}^{-1}$, compared to the neutron diffraction data.

The deviations from the neutron diffraction $F(Q)$ of the three classical systems, F3C, SPC/E and TIP3P are minimal and suggest that the simulation is sufficient in reproducing the structure of the system. However, gas-phase $F(Q)$ provides very different structural features compared to experiment.

It has been suggested that some of the observed discrepancies between simulation and experimental $F(Q)$ may be partly explained by the electronic structure and quantum nuclear effects [106]. In particular hydrogen is a very light atom and exists in abundance in this system. Very light atoms effect the structure factor directly due to their 'nuclear quantum effects', as discussed in refs [106] and [131]. The work by Troitzsch *et al.* [106] investigates the use of quantum



(b) $g_{O_{CE}-H_X}$)

Figure 3.3: Structure factor, F(Q), from Neutron Diffraction [3], the three glutamate aqueous solutions simulated used F3C, SPCE and TIP3P water potentials, and the gaseous simulation of a single glutamate molecule and sodium ion. The gas simulation has been offset by +0.15 for clarity.

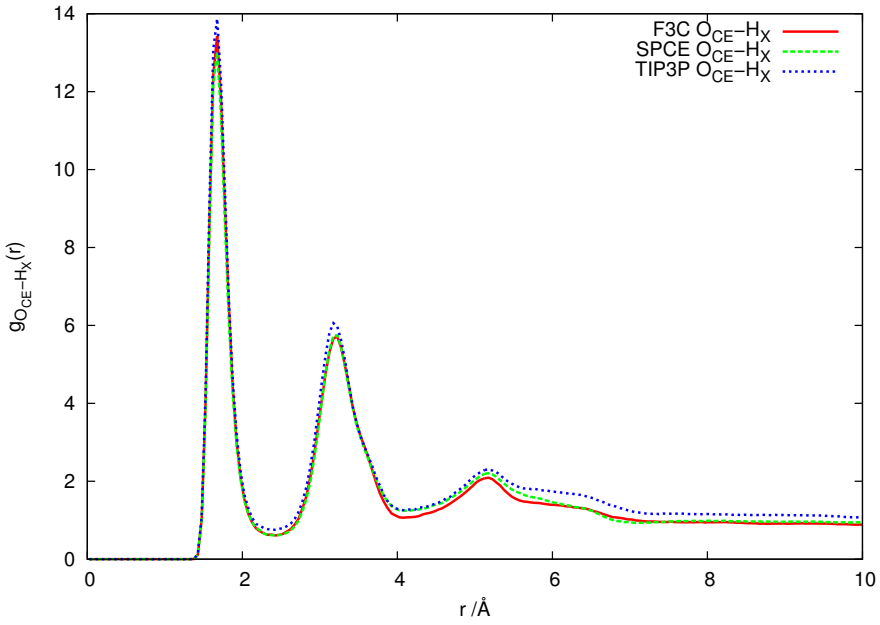
nuclear corrections finding that the corrections have the greatest impact to the structure factor in the high-Q region. Quantum corrections are not considered in this work as the effect on the low-Q peaks is less significant, and this is where the greatest discrepancies between the experimental and simulation data occurs. We propose that the under-estimation of the peaks in the $F(Q)$ in the simulated data may come from a combination of poor overall sampling of the system due to the shorter time-scales available to experiment

On the basis of this comparison we conclude that all three water models, in conjunction with the CHARMM22 empirical potential for glu and sodium ions, provides a description of the solution structure that will enable useful detail to be extracted from the simulations.

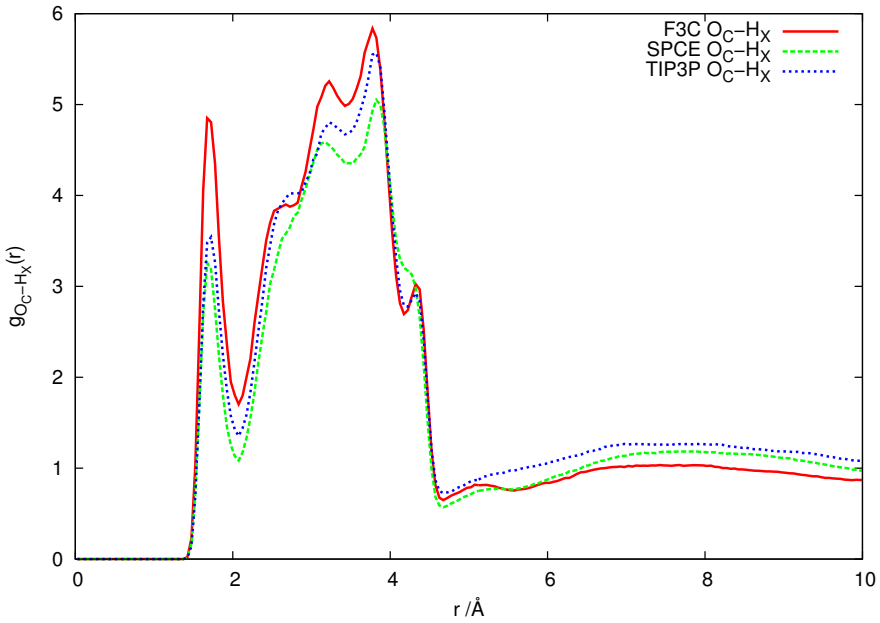
3.3.2 Glutamate-glutamate interactions

Molecules in solution may gain or loose protons depending on their pKa value and the pH of the solution that it is in. When the calculated pKa is equal to the solutions pH half of the molecules in the solution will loose protons (i.e. for a protonated molecule AH and a de-protonated molecule A^-, $\frac{A^-}{AH} = 10^{\text{pH-pKa}}$). Therefore, when a molecules' $pH > pKa$, there will be a higher proportion of molecules existing in a de-protonated than protonated form. As glutamic acid has a pKa value of 4.1, at pH7 the amino group, NH_3^+, the C_α-carboxyl of the backbone, $C_C - O_{C2}^-$, and the side-chain carboxyl, $C_{CE} - O_{CE2}^-$, are likely to all be ionised, and it is therefore a realistic assumption to consider only the fully deprotonated zwitterionic form of glutamic acid, glutamate in these simulations, with a net charge of -1e.. A labelled diagram of the molecule is provided in figure 3.1). Glu molecules form inter-molecular hydrogen bonds between the amine hydrogens (H_X) and the carboxyl oxygens (O_C and O_{CE}). Figure 3.4 shows the RDFs between the amine hydrogens and each of the carboxyl oxygens, indicating glu-glu hydrogen bond interactions. (It should be noted that these RDFs include inter and intra-molecular interactions.) There are clear differences, present in all three systems and are therefore not an artefact of the water potential, between the structures shown in the RDFs for each type carboxyl.

The first peak in each RDF describes the hydrogen-bond interaction and therefore the first-minimum indicates a radial cut-off for the hydrogen bond. Figure 3.4 shows well-defined first peaks for both carboxyl groups, indicating the



(a) $g_{O_{CE}-H_X}(r)$



(b) $g_{O_C-H_X}(r)$

Figure 3.4: RDFs for the side-chain carboxyl (3.4(a)) and the C_α-carboxyl (3.4(b)) with the amine hydrogen H_X.

existence of an $O \cdots H_X - N$ hydrogen bond, with a bond length of less than ~ 2.5Å. The intensity of the first peak for the side-chain-carboxyl RDF is nearly three times greater than that of the C_α-carboxyl RDF. The difference in intensity is highlighted by the co-ordination numbers shown in table 3.2. On average, the nitrogen (N) atom co-ordinates with just over one O_{CE} atom but only 0.3 O_C atoms in the F3C solution, with a similar trend observed for the other two water potentials considered. This indicates a significant difference in the number of bonds formed, depending on the species of carboxyl involved, with a preference to form hydrogen bonds between the side-chain carboxyl and the amine.

RDF	F3C		SPCE		TIP3P	
	r_{min} (Å)	$n_\alpha^\beta(r)$	r_{min} (Å)	$n_\alpha^\beta(r)$	r_{min} (Å)	$n_\alpha^\beta(r)$
$g_{H_X-O_{CE}}(r)$	2.43	0.32	2.38	0.31	2.38	0.35
$g_{H_X-O_C}(r)$	2.08	0.21	2.08	0.14	2.08	0.16
$g_{N-O_{CE}}(r)$	3.53	1.07	3.53	1.05	3.48	1.15
$g_{N-O_C}(r)$	3.78	0.33	3.73	0.21	3.83	0.28

Table 3.2: Hydrogen bond cut-off distance r_{min} and the corresponding co-ordination numbers $n_\alpha^\beta(r)$ for the the amine group with the two different carboxyl oxygens. Coordination numbers are for inter-molecular interactions (excluding intra-molecular) only.

The structure above 2 Å for the carboxyl-amine bonds is not due to a direct hydrogen-bond contact. Between 2.4 Å and 5 Å there is a great deal of discrepancy between the two carboxylate bonding structures. The side chain, $g_{O_{CE}-H_X}(r)$ shows a single peak in this region, corresponding to the second oxygen atom in the carboxylate coordinating with the same hydrogen. The C_α-carboxyl RDF, $g_{O_C-H_X}(r)$, forms a complex structure in this region. One broad peak encompasses small fluctuations, although it is of a similar intensity to the single peak of $g_{O_{CE}-H_X}(r)$. The third peak at $\sim$ 5Å is present in both RDFs, but is less clearly defined in $g_{O_C-H_X}(r)$. The third peak is consistent with the distance between the carboxyl oxygen and nitrogen atom a full molecule apart, suggesting a trimeric chain of glu molecules. It should be noted that such trimeric interactions are observed in an analysis of glu-cluster sizes, which are described in more detail at the end of this section.

The complex nature of the second peak in the $g_{O_C-H_X}(r)$ can be explained if the RDF *excluding* intra-molecular interactions is considered, as shown in

figure 3.5. This no longer exhibits the complex multiple peak structure observed in figure 3.4(b). Instead there is a simpler single second peak, as observed in the side-chain carboxyl interaction with the H_X atoms. This indicates that the multiple peak structure in the original RDF is due to the interaction of the three H_X atoms within the same molecule as the O_C atom. The H_X atoms are restricted in their freedom of movement with respect to the O_C atoms in the same molecule due to the constraints imposed on the backbone of the amino acid, explaining the three clearly defined peaks between 2 and 4.5 Å.

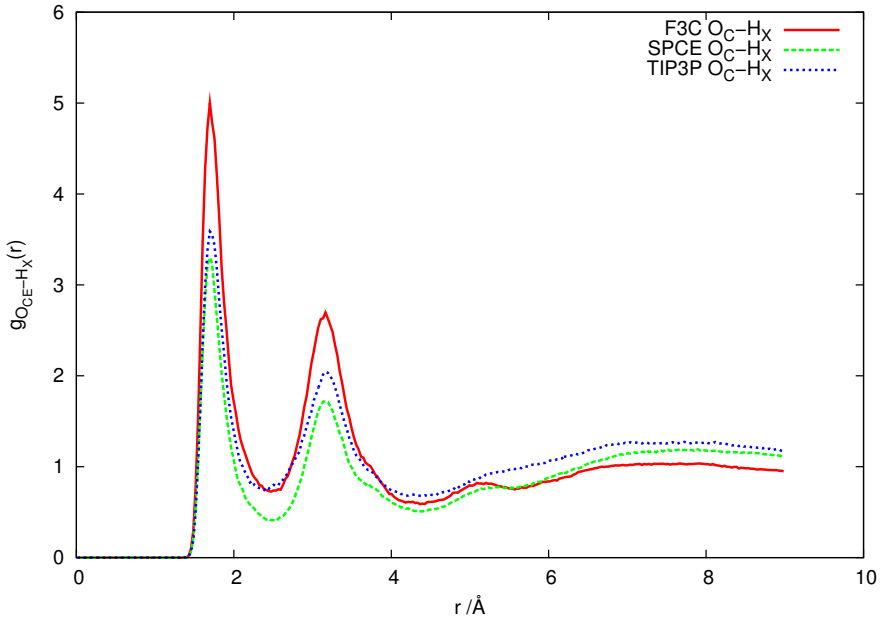
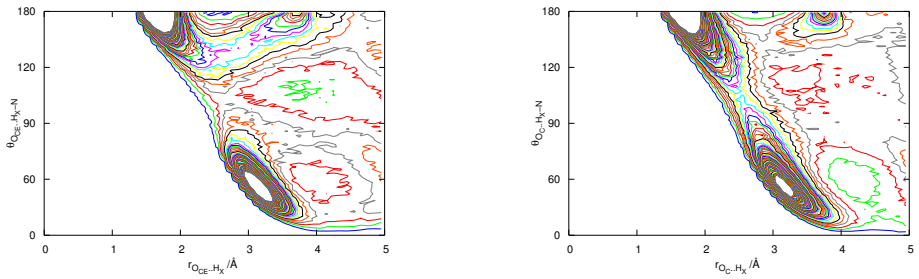


Figure 3.5: RDFs for $g_{O_C-H_X}(r)$ (C_α carboxyl) interactions, excluding all intramolecular interactions.

The removal of the multiple peak also shows that the peaks in the $O_C - H_X$ RDF in figure 3.5 are at similar radial positions to those of the $O_{CE} - H_X$ RDF in figure 3.4(a). The intensity of the first and second peaks in the side-chain-carboxyl RDF are nearly three times greater than the first and second peaks in the C_α RDF.

To further understand the $O_C \cdots H_X - N$ bonding interactions the 2D angular-radial and radial-radial distribution plots are considered. Figure 3.6 shows the correlation between the $O - H_X$ distance and the hydrogen bond angle, $\theta(O - H_X - N)$, for the two types of carboxyl oxygens in the glu solution with F3C water. A schematic of this type of contact is shown in chapter 2, figure 2.3(b).



(a) F3C $O_{CE}\cdot\cdot H_X - N$ (b) F3C $O_C\cdot\cdot H_X - N$

Figure 3.6: 2D angular-radial distribution function between each of the two carboxylate oxygens, O_{CE} (3.6(a)) and O_C (3.6(b)) atoms on glutamic acid and the amine group, NH_3^+, for glutamic acid in F3C water.

The angle shown is the linearity of the interaction, i.e. a fully linear topography will have a 180° angle. (It should be noted that in the definition of a hydrogen bond in chapter 2 that a linear interaction would be 0°.) The plots for the other two systems, SPC/E and TIP3P, show similar features and are shown in appendix A, figure A.1.

There are two clear structural features in the plots for both the O_C and the O_{CE} carboxyls. The first at $r_{O-H_X} \sim 1.6$ Å, $\theta \sim$ 180° is due to the short, strong, linear hydrogen-bond contact between the two groups. This linear bond feature is narrower in angular terms for the O_{CE} atoms, suggesting a stronger hydrogen bond with less angular flexibility than in the C_α-carboxyl-amine interaction.

The wider spread of angles in the short-range contact ($r_{O_C-H_X} \sim$ 1.6 Å) enables the formation of bifurcated bonds, where one hydrogen is shared between two oxygen atoms. The greater angular spread of the O_C contact at $\sim$ 1.6 Å indicates a greater propensity to form bifurcated bonds *via* the C_α-carboxyl, than the side-chain-carboxyl. The second contact is at a larger radius with $r_{O-H_X} \sim 3.2$ Å and $\theta \sim$ 60°. The angle and length of this contact is consistent with two motifs. Firstly, the co-ordination of the amine H_X atom with the unbonded oxygen atom from the carboxyl, and secondly, when the bonded oxygen co-ordinates with an unbonded H_X atom from the same amine group.

The existence of a bifurcated bond, where one atom is shared between two other atoms, as illustrated by the trajectory snapshot in figure 3.7, is most clearly identified by 2D radial-radial distribution plots. Figure 3.8 shows the 2D

radial-radial distributions correlating the radial distance between the two types of carboxyl oxygen (O_{CE} and O_C) and the amine-hydrogen (H_X). A schematic of this type of correlation is shown in chapter 2, figure 2.3(a). Again, only the plots for the glu molecules in the F3C system are shown. Similar plots were obtained for the SPC/E and TIP3P simulations, and these are shown in appendix A, figure A.2.

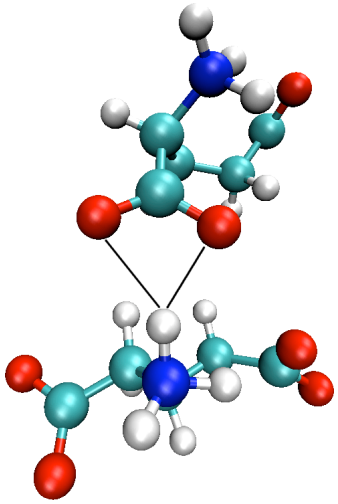
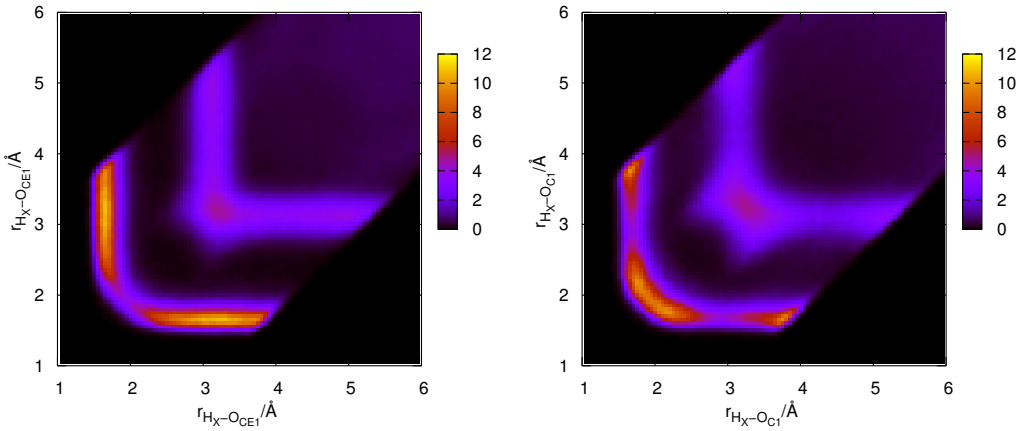


Figure 3.7: Bifurcated hydrogen bond between NH_3^+ group and C_α carboxylate group ($C_C - O_C 2^-$).

Both the C_α and the side-chain 2D radial-radial distributions show a sharp intensity at $r \sim 1.7$ Å for both r_1 and r_2. This peak can be explained by two different molecular configurations. In the first motif both radii are equal, $r_{O_1-H_X} \simeq r_{O_2-H_X} \sim 1.7$ Å, indicating a bifurcated bond, which is highlighted by the region of intensity in this region on the plots. The C_α-carboxyl has a greater propensity to form these bifurcated bond-states than the side-chain carboxyl. The second motif, where r_{O-H_X} is up to 3.9Å, is consistent with a single linear hydrogen bond (the peak at $r_{O-H_X} \sim 1.6$ Å, $\theta \sim 180°$ in the radial-angular distributions). The small radial distance is due to the hydrogen-bond, and the larger radial interaction is due to the second carboxyl oxygen, not directly bonded to the amine, interacting with the H_X atoms.

The C_α radial-angular distribution in figure 3.6 has a broader angular spread for the two main peaks than the side-chain plot. The broader angular spread is caused by a larger range of angles for the $O_C - H_X$-N hydrogen bond being



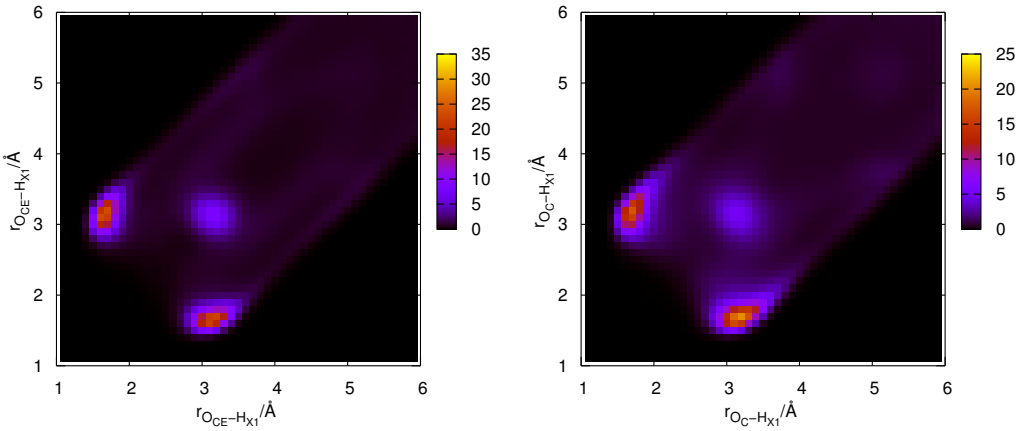
(a) F3C $H_X \cdot\cdot O_{CE}$ (b) F3C $H_X \cdot\cdot O_C$

Figure 3.8: 2D radial-radial distribution between the two species of carboxylate oxygens, O_{CE} and O_C, and the amine hydrogens (H_X), for glutamate in the F3C solution.

stable than in the side-chain case. A broader angular spread is consistent with the presence of bifurcated bonds, as this enables hydrogen-bonds to form and remain stable at angles up to 60° away from the linear configuration, compared with 30° in the standard definition of a hydrogen bond as described in Chapter 2. The greater propensity of the C_α-carboxyl group to form bifurcated bonds, as shown by the 2D-radial-radial distribution plots (figure 3.8), is consistent with the greater spread of angles seen in the 2D angular-radial distribution plots in figure 3.6(b).

Figure 3.9 shows the 2D radial-radial distribution plots correlating amine-carboxyl radial distances, $H_{X1} - O$ and $H_{X2} - O$, where H_{X1} and H_{X2} are two hydrogens within the same amine group. If a peak exists in the region $r_{H_{X1}-O} \simeq r_{H_{X2}-O} \sim 1.7$ Å, this would indicate the existence of bifurcated bonds involving two amine hydrogens bonding to a single carboxyl-oxygen. Both plots show zero intensity in this region, indicating that bifurcated bonds of this type do not form *via* either the side-chain or C_α-carboxyl. The plots for the other two systems, appendix A, figure A.3, show similar results.

On observation of the trajectory, many of the bifurcated bonds appeared to exist as a transition state between two linear bonds. This is where atom A is bonded *via* a single linear bond to atom C, followed by the formation of a bifurcated state where atoms A and B are both bonded to atom C, and then the first atom moves away leaving a linear bond between B and C. This may



(a) F3C $O_{CE} \cdots H_X$ (b) F3C $O_C \cdots H_X$

Figure 3.9: 2D radial-radial distribution functions for the glu-glu amine-carboxyl interactions correlating the $r_{O_{CE}/O_{C1} \cdots H_{X1}}$ with $r_{O_{CE}/O_{C1} \cdots H_{X2}}$ in solutions of CHARMM22 Glutamate with F3C water.

be because the bifurcated state provides a lower energy pathway between two linear bonds compared to the energy required to completely break and reform such bonds. The low energy pathway is consistent with shorter lived bonds, and providing a possible explanation for the smaller proportion of glu-glu bonds forming *via* the C_α-carboxyl, which forms proportionally more bifurcated bonds, compared to the side-chain. This is also suggested by Leenders *et al.* [44] in their *ab initio* study of glu bonds with water.

There may be several contributing factors producing the different behaviour of the two carboxylate groups. Firstly, the charge on the two different species of oxygen atoms is different. O_C has a charge of -0.67 e, O_{CE} a charge of -0.76 e in the CHARMM22 force-field. The greater charge of the O_{CE} atoms is consistent with stronger, longer lived bonds, as the attraction between the oxygens and the positively charged H_X atoms is greater, and is also consistent with a larger co-ordination number for the $O_{CE} - H_X$ interaction. This charge differential may also contribute to the greater angular spread seen in the peaks on the 2D radial-angular distribution functions for $O_C - H_X$ compared to $O_{CE} - H_X$, and also the differing proportions of bifurcated bonds. O_{CE} may bifurcate less as its linear hydrogen bonds are stronger, therefore not requiring the bifurcation to stabilise the bond. The charge difference is a characteristic of not just the CHARMM22 model but also of other empirical potential models for amino acids. For instance, although the Amber force-field [132] does not explicitly treat single

amino acids and does not satisfactorily parameterise glutamate, information on the termination of the charged groups enabled Daub *et al.* to average the two terminal cases and then scale the charges to obtain an overall charge on the molecule of -1 e. This will have resulted in charges for O_C and O_{CE} that are significantly different from each other, as in the CHARMM22 parameterisation used in this study.

Secondly, there may also be steric and electrostatic impediments reducing the number of amine-C_α bonds. The $-NH_3^+$ group may prefer forming bonds with the side-chain carboxyl as bonding here maximises the distance between the amines in two adjacent glu molecules. If bonding occurs through the O_C atoms alone, then two amine groups are less than 4.5 Å apart, providing little shielding between the positive charges. The side-chain carboxyl is further away from the amine, increasing the amine-amine distance. Conformational flexibility may also play an important role. The side chain can bend, enabling the distance between two amines to be further maximised, whereas the amine and C_α-carboxyl are in a more tightly constrained structure. The side-chain movement also enables the O_{CE} carboxyl to move towards a bond site on an adjacent molecule more freely than the rigidly fixed O_C carboxyl which requires the entire amine/O_C carboxyl group to be displaced. This is particularly important in this system as the high concentration of glu means that molecular movement is likely to be inhibited by the presence of other glu molecules.

Figure 3.10 shows the distribution of glu cluster sizes averaged over the entire production run. The distribution gives the fraction of clusters of a particular size, obtained from instantaneous trajectory snapshots, as a proportion of all the clusters over the entire simulation, including clusters of just a single member. No percolating clusters, containing the possible maximum of 20 molecules, were found in any of the systems throughout the entire trajectory. The majority of clusters found were dimers, trimers or tetramers, and a large proportion of molecules form no glu-glu bonds.

All three of the aqueous solution simulations exhibit the same trend, with the TIP3P system forming the largest clusters, although the proportion of clusters between 3 and 6 members falls off more rapidly for the TIP3P system than the F3C and SPC/E systems (note the log-scale in figure 3.10). This is consistent with a significantly smaller peak seen in the $O_C - N$ RDF (not shown here) for

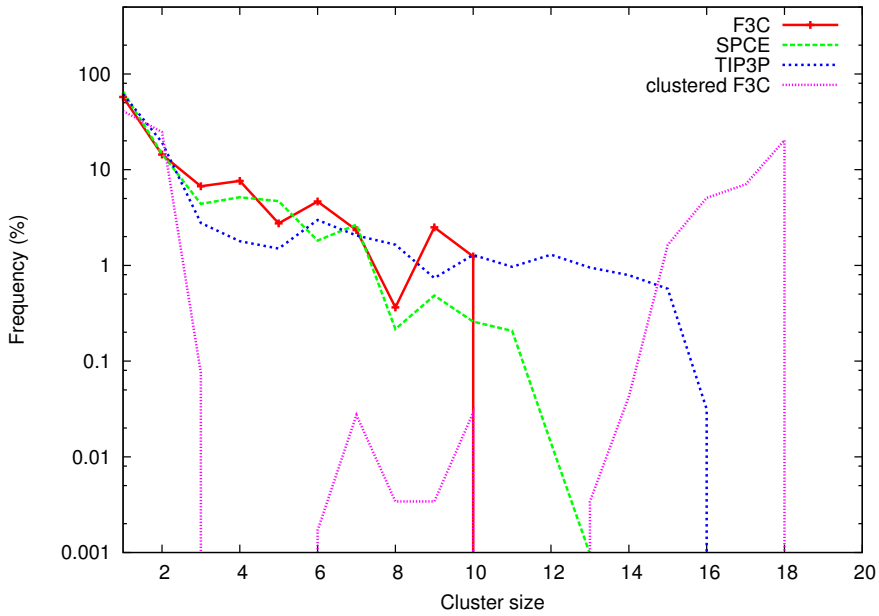


Figure 3.10: Histograms of fractions of the size of clusters found in the simulation as a function of the total number of clusters..

the TIP3P system, suggesting a smaller number of $O_C - N$ bonds, and therefore proportionally fewer larger clusters compared to SPC/E and F3C water solutions.

The small cluster sizes are consistent with the bonding between C_α carboxyls and amines being restricted, as larger cluster formations are inhibited due to the orientation of the molecules. This, combined with the smaller number of $O_C - H_X$ bonds would suggest that the C_α-carboxyl would have greater hydration than the side chain. Table 3.3 gives the co-ordination numbers of water around the two types of carboxyl oxygens. This shows a greater co-ordination number for water around the side-chain oxygen, O_{CE}, compared to the C_α-carboxyl-oxygen, O_C for the F3C and SPC/E water solutions. This is in agreement with the simulation results reported by Daub *et al.* [45]. The trend is reversed in the TIP3P system, but this is due to the radial cut-off distance being large due to the flattened structure of the corresponding RDF pushing the cut-off distance to larger r.

To ensure that the clustering observed here is not an artefact of the initial configuration another simulation was run with identical simulation parameters but using an initial starting configuration consisting of aggregated glu molecules. Several water molecules were within the aggregate to mediate energetically-unfavourable glu-glu interactions, with the rest of the water molecules and the

RDF	F3C		SPCE		TIP3P	
	r_{min} (Å)	$n_\alpha^\beta(r)$	r_{min} (Å)	$n_\alpha^\beta(r)$	r_{min} (Å)	$n_\alpha^\beta(r)$
$g_{O_C-O_W}(r)$	3.13	2.19	3.03	2.02	3.73	4.33
$g_{O_{CE}-O_W}(r)$	3.18	2.62	3.13	2.52	3.22	2.61
$g_{O_C-H_W}(r)$	2.43	1.97	2.38	1.96	2.43	2.01
$g_{O_{CE}-H_W}(r)$	2.38	2.44	2.33	2.36	2.43	2.28
$g_{N-O_W}(r)$	3.48	2.59	3.38	2.57	3.48	2.73
$g_{H_X-O_W}(r)$	2.38	0.59	2.38	0.65	2.33	0.59

Table 3.3: Hydrogen bond cut-off distance r_{min} and the corresponding co-ordination numbers $n_\alpha^\beta(r)$ for the the two the two carboxyl oxygens, the C_α-carboxyl, O_C, and the side-chain carboxyl, O_{CE}, and the amine-nitrogen, N, with water oxygens (O_W) and hydrogens (H_W). The data provided is as in table 3.2.

sodium ions surrounding the large cluster. The clustering data from this run is shown by the dotted purple line in figure 3.10. In this 'clustered' simulation an initial separation of the aggregated glu molecules was observed following by a swift re-clustering into one large group, a percolating cluster, such that they were all connected to one another with glu-glu hydrogen bonds. The separation and subsequent re-formation of the cluster appeared to enable water molecules to enter the aggregate to provide further mediation of undesirable glu-glu interactions. This aggregate then remained stable for the entirety of the 1 ns simulation, however, one or two members often broke away temporarily from the large cluster, resulting in a maximum cluster size of 18 rather the full 20 glu molecules available.

These results highlight the importance of the initial starting configuration for this system, and suggest very slow dynamics within the solution. To ensure that our other results discussed here were still relevant the RDFs for the clustered system were compared with the original production run RDFs. Both sets of RDFs shared the main structural features, although there was some fluctuation in the intensity of the peaks, but the difference in co-ordination numbers was not significant. This indicates that despite the dependence of initial configuration the structural motifs identified in this study are not unique to a single starting configuration. This presumably affects all other reported simulations on this system [3, 44, 45]. In the subsequent chapter this is examined further with the use of parallel tempering [97]. Parallel tempering enables enhanced sampling of the potential energy surface of systems such as these and therefore is an ideal

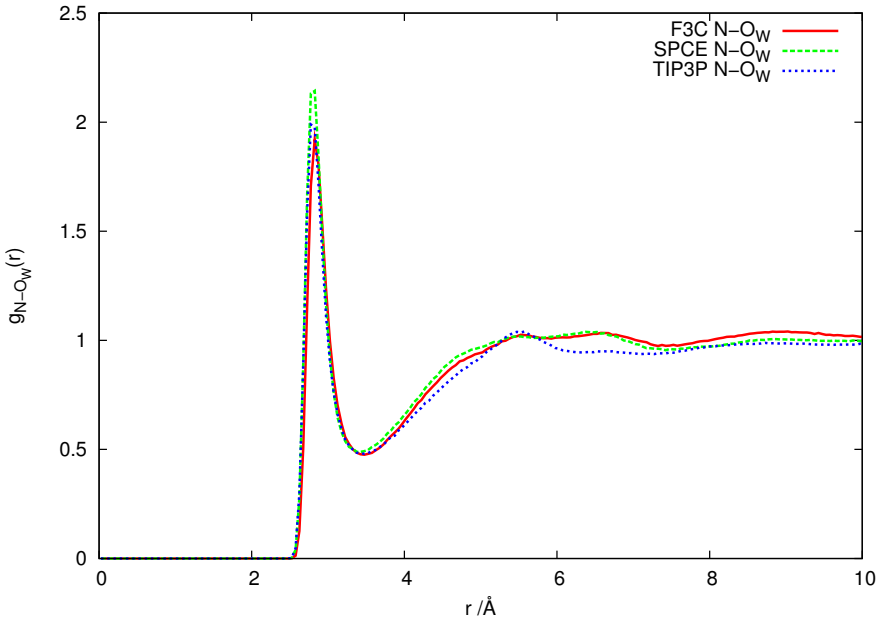


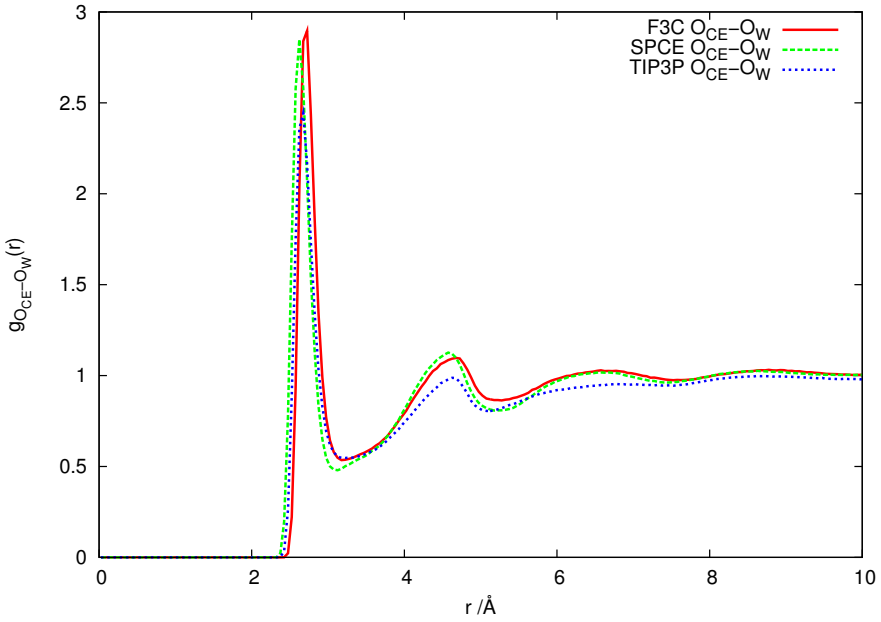
Figure 3.11: Glu-water RDFs for the $N - O_W$ interaction.

method to apply in this situation.

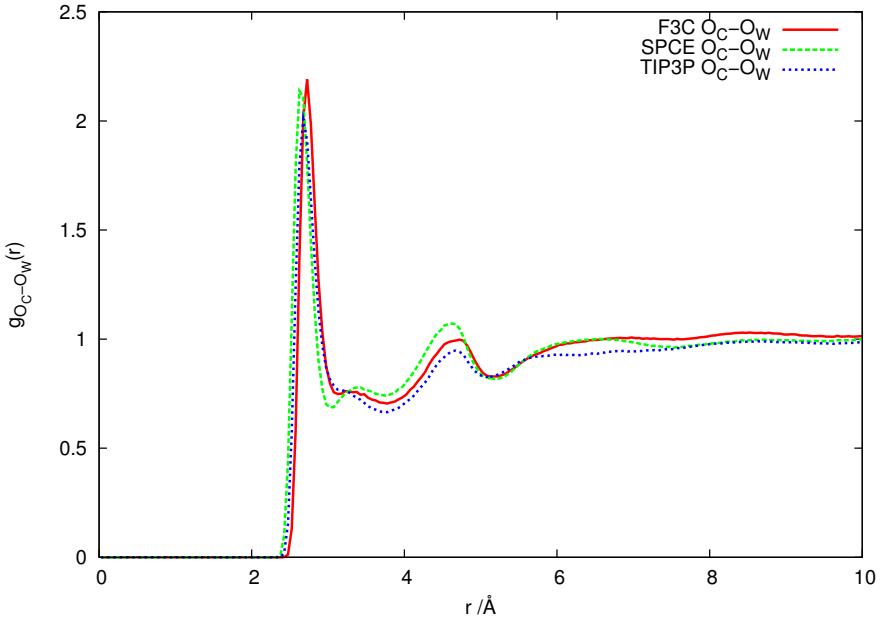
3.3.3 Water and ion interactions

To further investigate the different behaviour between the charged groups of the glu molecule water-glu interactions are now considered. Figures 3.11 and 3.12 show the amine-water and carboxyl-oxygen-water radial distribution functions. All three groups show clearly defined first peak in the region $r < 3.5$ Å, consistent with the hydrogen-bond interaction.

The two carboxyl group oxygens form between 2 and 3 hydrogen bonds with water (see co-ordination numbers in table 3.3). The nitrogen acts as a proton donor with H_X atoms on the amine bonding to the oxygen atoms of water, with a coordination of 0.59 atoms per amine group. This suggests that not all H_X atoms hydrogen bond with water, indicating that most amine groups are involved in glu-glu bonds. All three water models provide a similar description of the solvation structure. The RDF of the side-chain carboxyl has a greater intensity for the first peak than the C_α-carboxyl, as seen in the corresponding co-ordination numbers (table 3.3). As mentioned in the previous section, this is mirrored by the corresponding glu-glu interactions where the $O_{CE} - H_X$ RDF had a first peak



(a) $g_{O_{CE}-O_W}(r)$



(b) $g_{O_C-O_W}(r)$

Figure 3.12: Glu-water RDFs for the $O_{CE} - O_W$ and $O_C - O_W$ interactions.

nearly three times that of the $O_C - H_X$ RDF. This is also in agreement with the work by Leenders *et al.* [44] in the simulation of a single glu in water, suggesting that this interaction is not dependent on other glu molecules in the solution. In each case a second peak exists in the water-glu RDF. In the $N - O_W$ RDF this corresponds with a second hydration shell, although the peak is very poorly defined. In the case of the carboxyl-O_W RDFs the second peak corresponds to the second oxygen atom in the carboxyl, where the first oxygen is bound to the water molecules resulting in the first peak of the RDF. There is a suggestion of a third peak in the carboxyl-O_W RDFs, corresponding to a second, poorly defined, hydration shell. None of the three charged groups of the glu-molecule perturb the water structure enough to form a clear second hydration shell.

As seen in the glu-glu interactions there is a difference in the bonding behaviour of the two carboxyl groups. The C_α-oxygen RDF shows a shoulder on the tail of the first peak, which is not seen for the side-chain carboxyl-oxygen. This discrepancy is present in all three systems and and so is not an artefact of the water model. Daub *et al.* [45] identify a similar shoulder for the C_α-oxygen - water-oxygen RDF of the classical glu solution at a concentration of 1:1:36 (glu:sodium:water), although the shoulder is on the leading edge of the second peak. Their work did not find this peak at other concentrations or in the *ab initio* calculation where a single glu molecule was in solution with water but no sodium ions. This suggests that this motif may be caused by the presence of glu-glu interactions, not present in solutions containing only one glu molecule, or because of the presence of the sodium ion.

Daub *et al.* attribute this shoulder on the RDF to a sodium ion mediating the carboxyl-water interaction, such that the carboxyl bonds to the sodium ion, which in turn bonds to a water molecule (as shown in the illustration in figure 3.13), producing $O_C - O_W$ distances in the range of the intermediate peak. The shoulder in our data is produced by a motif where the radial distance between the O_C and O_W atoms is between 3.1 Å and 3.8 Å. On observation of the trajectory, using VMD, water molecules and sodium ions were found to be closely bound to the same carboxyl oxygen. This motif was found on both the side-chain and the C_α-carboxyl groups and therefore this does not explain the difference between the two RDFs. In addition, on further analysis of the trajectory, when $r_{O_{CE}-Na^+} <$ 3.13 Å and $r_{O_W-Na^+} <$ 3.08 Å (for the same sodium ion) for the same O_C and

O_W atoms the $r_{O_C-O_W}$ distance was *never* in the range between 3.1 and 3.8 Å as required for the intermediate peak throughout the entire trajectory. Therefore, this peak cannot be due to sodium ions mediating the $O_C - O_W$ interaction as suggested by Daub *et al.*.

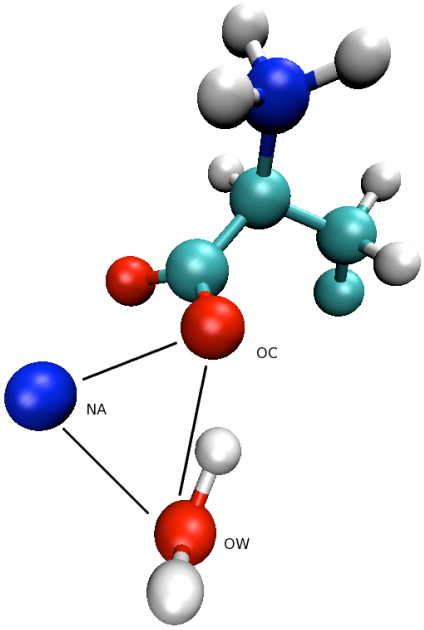


Figure 3.13: Diagram of the O_W atom in water (bottom) coordinating with the glu-C_α-carboxyl oxygen (OC - top), mediated by a sodium ion (Na^+ -left).

The O-O distance between 3.1 Å and 3.8 Å is consistent with the breaking and formation stages of a bifurcated bond between the carboxyl-oxygens and water. A bifurcated bond between water and the carboxylate is formed by sharing the water-hydrogen, H_W between the two carboxyl oxygens. On visualisation of the trajectory bifurcated bonds were often formed by initially forming a single water-glu bond before sharing the H_W atom on the water molecule between the two carboxyl oxygens. This movement, and the subsequent breaking move, provides $O_W - O$(carboxyl) distances of between 3.1 Å and 3.8 Å, consistent with the shoulder in the RDF. Bifurcated bonds exist in a similar propensity between the carboxyl oxygens and water for both the O_C and O_{CE} atoms. This motif does not adequately explain the shoulder on the $O_C - O_W$ RDF.

An alternative explanation for the shoulder on the C_α-oxygen-water RDF is when a water molecule is bonded to both the amine and C_α-carboxyl, producing a 'circular' C_α-carboxyl-water-amine motif, with a slightly stretched $O_C - O_W$

bond-length. A schematic of this is shown is figure 3.14. This motif can only occur with the O_C atoms, as the O_{CE} atoms are too far away from the amine, and therefore could explain the difference in the RDFs. To identify if this motif is present 2D radial-radial distributions were calculated correlating the $O_C - O_W$ and the $N - O_W$ distances (figure 3.15). (The plots for the SPC/E and TIP3P systems are shown in appendix A, figure A.4.) This plot shows a non-zero intensity in the region $r_{O_C - O_W} \simeq r_{O_W - N}$, suggesting that this motif may be present.

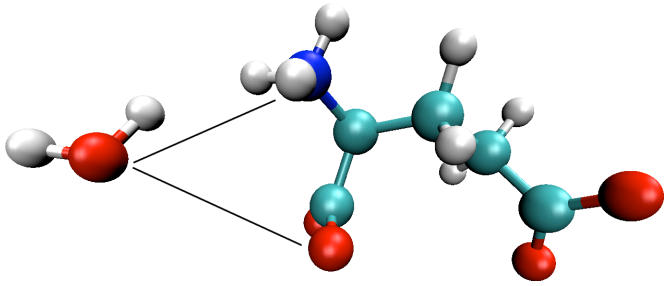


Figure 3.14: Image of water molecule shared between glutamate C_α-caboxyl and amine groups, with the $H_W - O_C$ and $O_W - H_X$ distances highlighted.

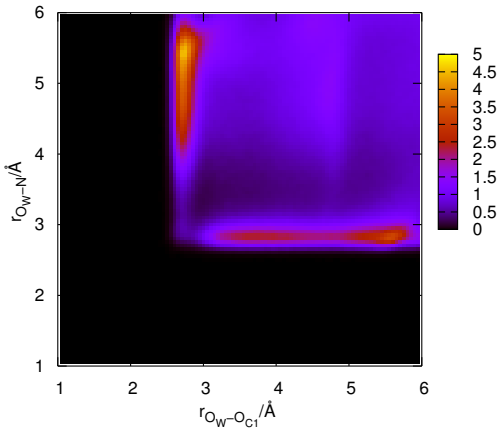


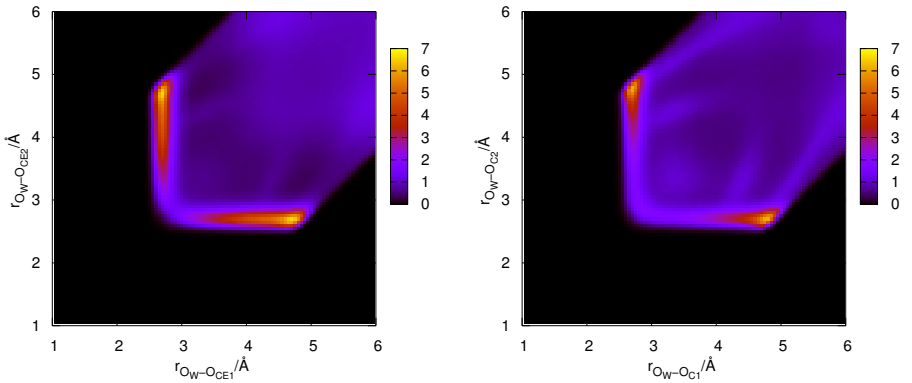
Figure 3.15: 2D radial-radial distribution correlating the $O_W - O_C$ distance with the $O_W - N$ distance such that the O_C and N atoms are within the same molecule (as shown in the illustration in figure 3.14). This is for the F3C system, with similar results obtained for SPC/E and TIP3P.

Further analysis of the trajectory then enabled us to identify the proportion of glu-water bonds that were involved in this 'circle'-motif. If a hydrogen-bond was defined by just the oxygen-oxygen ($O_C - O_W$) and oxygen-nitrogen ($O_W - N$) bonding distances, then ~8.4% of the total $O_W - N$ hydrogen bonds and ~0.4%

of the total $O_C - O_W$ hydrogen bonds were involved in the formation of a circular motif. Refining the hydrogen bond definition to include an $O_C - H_W$ and $H_X - O_W$ cut-off distance meant that no circular motifs of this nature were found. Instead, the water molecule was found to be hydrogen bonded to the amine group of the glu, such that the water-oxygen (O_W) is within ~ 3.1 Å of the carboxyl oxygens (O_C). Both the water-hydrogens (H_W) are orientated away from the oxygen and no hydrogen bond is actually formed between the O_C and water. This suggests that to produce a $N - H_X \cdots O_W - H_W \cdots O_C$ circular motif requires the H_X and H_W hydrogens to be closer than their electrostatic repulsion allows. Although the water molecule is not hydrogen bonded to the carboxyl oxygen, the O_W is held in this close proximity to the O_C atom and therefore may be contributing towards the shoulder on the $O_C - O_W$ RDF. The circular motif of C_α-carboxyl-water-amine may contribute to the $O_C - O_W$ RDF intermediate peak, although the C_α-carboxyl-water relationship is not hydrogen bonded.

We found that the circular motif accounted for $\sim$2% of the total $O_C - O_W$ bonds in the region $3.1 < r < 3.8$Å. It would be interesting to calculate the percentage of the $O_C - O_W$ bonds in the region $3.1 < r < 3.8$Å that contribute to the intermediate peak of the RDF. However, we have not found a robust method of calculating the contribution to the total number of $O_C - O_W$ bonds, and therefore cannot determine accurately if the circular motif accounts for all of the interactions in the region of the intermediate peak that differentiates the O_C and O_{CE} RDFs.

Figure 3.16 shows the 2D radial-radial distribution plots between each of the carboxyl groups and water in the F3C system. Similar plots were obtained for the SPC/E and TIP3P systems, shown in appendix A, figure A.5. We see that the plots are similar for both carboxyls, each forming bifurcated bonds with water molecules. These results contrast with the glu-glu interactions where the C_α carboxyl oxygen preferred forming bifurcated bonds compared to the side-chain carboxyl. In addition Daub *et al.* [45] did not find a bifurcated bond motif between either species of carboxyl oxygen and water, although the EPSR data obtained by McLain *et al.* [3] did predict bifurcated motifs. To assess whether this is an artefact of potential and/or system size a system was simulated using the same parameters and empirical potentials used in the main discussion, but using a concentration of 1:1:36 and reducing the number of molecules to 6 glu,



(a) F3C $O_W \cdot\cdot O_C E$ (b) F3C $O_W \cdot\cdot O_C$

Figure 3.16: 2D radial-radial distribution between F3C water-oxygen (O_W) and each of two glu-oxygen species.

6 sodium ions and 216 waters as simulated by Daub *et al.* [45]. We found that the co-ordination numbers for our 1:1:29 simulation, Daub *et al*'s work, and our 1:1:36 simulation were comparable. We also found bifurcated motifs between water and the carboxyls in a similar propensity in the 1:1:36 simulation. This suggests that the absence of the bifurcated bonds is not due to the system size or concentration, but due to the empirical potential. It has previously been suggested that the CHARMM22 force-field over-estimates the presence of bifurcated motifs [114]. As the bifurcations are also present in the EPSR data, this does not necessarily explain the disparity between our results and Daub *et al.* [45]. The presence of a bifurcated-bond state may contribute to the $O_C - O_W$ RDF intermediate peak, but would then suggest a different formation mechanism to the $O_{CE} - O_W$ bifurcated bonds. In addition the amine-water 2D radial-radial distribution (not shown here) shows that the H_X hydrogens only ever form single linear bonds with water molecules, as found for the amine-carboxyl 2D radial-radial plots, indicating that bifurcated bonds on this site are not energetically favourable.

The bonding that occurs between the water molecules and glu will affect the bulk water structure. In figure 3.17 the water-glu hydrogen bonds are compared with water-water hydrogen bonding both in the glu systems and for a simulation of pure water. The main distinction between pure water and the water in the glu solution is a 'smearing' in the definition of the first trough of the $g_{O_W-O_W}$ RDF in the glu-system. This in turn pushes the position of the first minimum out to 3.725 Å compared to 3.375 Å, thereby increasing the coordination number

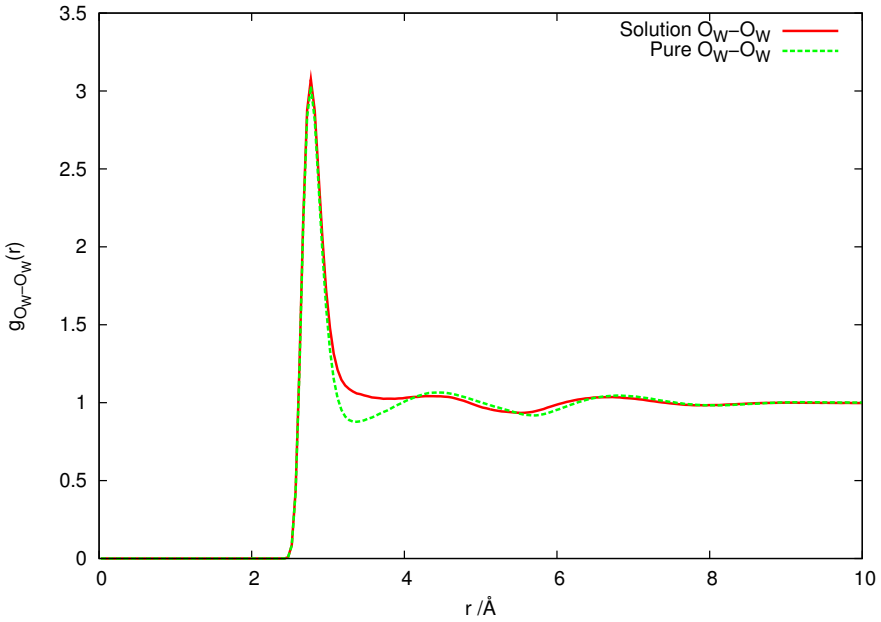


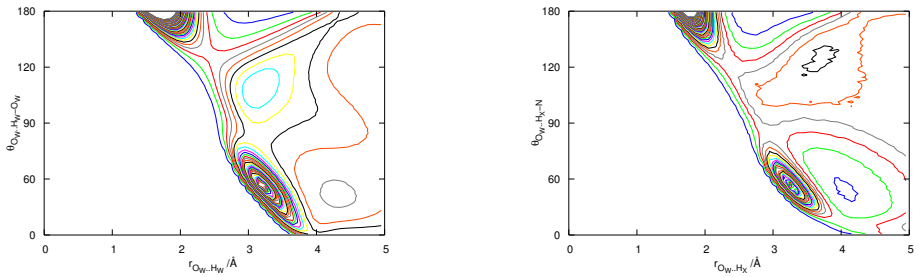
Figure 3.17: Radial distribution functions comparing F3C water-water (in pure water and glutamate solution water) interactions with the associated water-solute interactions.

(as seen in table 3.4). If the first minimum is taken as 3.375 Å (the same as in pure water), tthere is a clear reduction in the cooridnation number compared to the pure water RDF, from 4.7 to 4.4. This is consistent with the observation by McLain *et al.* [3] where the coordination number reduced from 4.5 to 4.3. There is also less perturbation of the first shell structure from pure water than the EPSR results obtained by McLain *et al.* [3]. In the artificially clustered system (discussed above), the water-water RDFs were more like pure water, with less 'smearing' of the first trough. This is consistent with fewer glu-water interactions, and no clear second hydration shell around the glu molecules.

Figure 3.18 shows the 2D radial-angular distributions for the water-water interaction, $O_W \cdots H_W - O_W$, and the water-glu interaction $O_W \cdots H_X - N$, correlating the $O_W - O_W/N$ radial distances with the angle of the bonds. These both produce similar plots with a peak at an O-H distance of $r \sim 2$ Å corresponding to a linear bond at 180°, and a peak at $r \sim 3$ Å, $\theta \sim 60°$ corresponding to the co-ordination of the second, unbonded, hydrogen atom. Neither show a distinct preference for a second hydration shell at 5Å, although the $O_W - O_W$ RDF would indicate that this is more likely in the water-water

RDF	F3C		SPCE		TIP3P	
	r_{min} (Å)	$n_\alpha^\beta(r)$	r_{min} (Å)	$n_\alpha^\beta(r)$	r_{min} (Å)	$n_\alpha^\beta(r)$
$g_{O_{CE}-Na^+}(r)$	3.13	0.67	3.08	0.74	3.18	0.76
$g_{O_{CE}-Na^+}(r)$	3.03	0.36	2.98	0.33	3.03	0.50
$g_{O_W-Na^+-O_C}(r)$	3.08	3.80	3.03	3.62	2.98	3.39
$g_{O_W-O_W}(r)$ pure	3.38	4.77	3.33	4.47	3.68	6.45
$g_{O_W-O_W}(r)$ soln	3.73	6.07	3.58	5.27	4.38	10.37

Table 3.4: Hydrogen bond cut-off distance r_{min} and the corresponding co-ordination numbers $n_\alpha^\beta(r)$ for sodium ions, Na^+, around glu and water oxygen atoms. Also shown is the $O_W - O_W$ data for a system of pure water.



(a) F3C $O_W \cdot\cdot H_W - O_W$ (b) F3C $O_W \cdot\cdot H_X - N$

Figure 3.18: Comparison of the 2D radial-angular distribution for a water-water hydrogen bond, $O_W \cdots H_W - O_W$, and for the glu-amine-water hydrogen bond, $O_W \cdots H_X - N$ (from the F3C solution, with similar results obtained for SPC/E and TIP3P).

interaction than the water-amine interaction. The same was found for the SPC/E and TIP3P solutions, as shown in appendix A, figure A.6.

Finally, sodium ion interactions with glu and water molecules are considered. Figure 3.19 shows the RDFs for sodium ions (Na^+) interacting with the three oxygen species in the system, O_{CE}, O_C and O_W. The strong first peak indicates sodium-oxygen bonds forming at $\simeq$ 2.2Å. The $Na^+ - O_C$ RDF has the greatest intensity first peak, indicating a greater proportion of sodium-C_α-carboxyl interactions, however the greatest $O - Na^+$ coordination number is with between sodium ions and water, although the $Na^+ - O_C$ is nearly double the $Na^+ - O_{CE}$ co-ordination number, indicating that glu-sodium bonds prefer to form *via* the C_α carboxyl. This may be due to the mobility of sodium, as

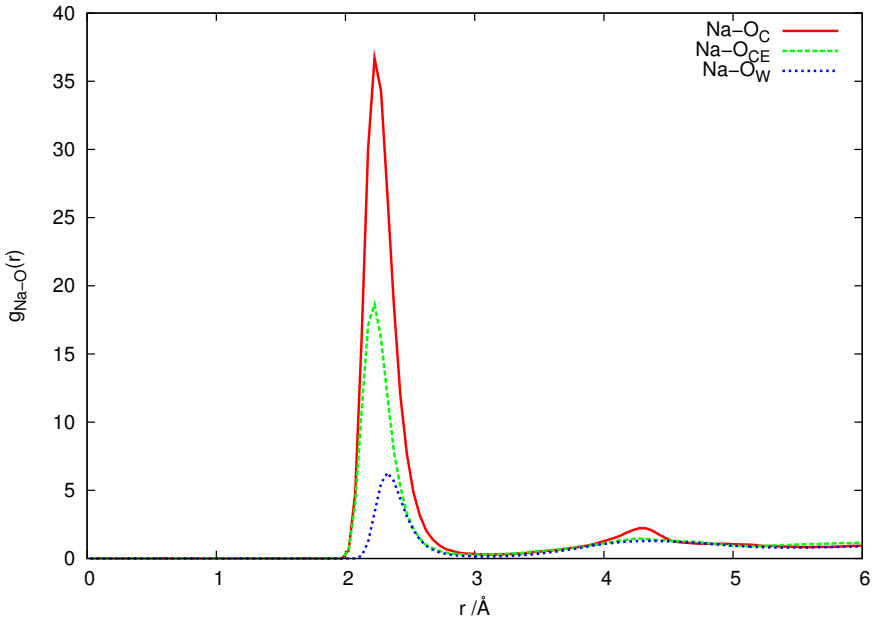


Figure 3.19: Radial distribution functions for sodium ions with the two carboxyl groups of glutamate (from the F3C solution, with similar results obtained for SPC/E and TIP3P).

a single ion, facilitating its ability to overcome the restricted movement of the C_α-carboxyl and the proximity of the electrostatic repulsion of the amine group.

Analysis of the $Na^+ - O_C$ and $Na^+ - O_{CE}$ bonds showed that 55% of the $Na^+ - O_C$ bonds and 50% of the $Na^+ - O_{CE}$ bonds formed were bifurcated, where the two oxygen atoms on the same carboxyl group both co-ordinated with the same sodium ion. The high proportion of bifurcated states may be an artefact of the empirical force-field, as mentioned previously.

Sodium ions fully hydrated with water have been found to have a co-ordination number of 6.53 in a classical Molecular Dynamics simulation of an aqueous NaCl solution [51]. In this system Na^+ is hydrated with an average of 3.8 water molecules (table 3.4). This could mean that either all sodium ions are hydrated with between 3 and 4 water molecules, or, alternatively there could be some sodium ions with a full complement of six water neighbours and then the rest with far fewer than four. A histogram showing the fraction of sodium ions hydrated by between 0 and 6 waters and the corresponding fraction of sodium ions hydrogen bonded to between 0 and 6 glu molecules is shown in figure 3.20.

We see that sodium ions are most commonly hydrated by four water

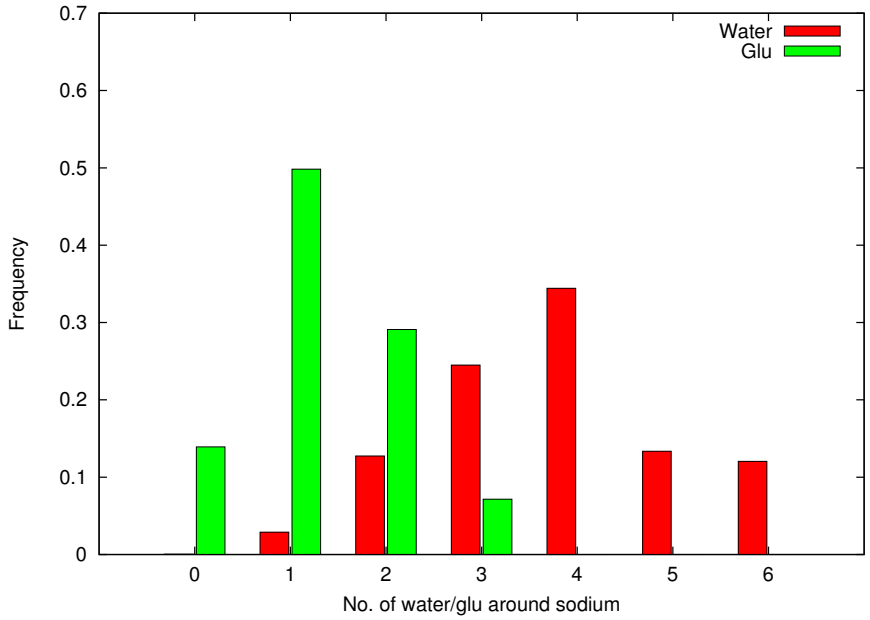


Figure 3.20: Histogram of the number of (F3C) water molecules (red) and the number of glutamate molecules (green) hydrating sodium ions as a fraction of the total number of sodium ions.

molecules, suggesting that most sodium ions are also bonded to at least one glu-molecule, however over 10% of the ions are fully hydrated, with no glu contacts. Correspondingly glu molecules never provide more than three direct sodium-glu contacts per sodium ion, with the majority of sodium ions forming two glu bonds, although some form up to three. The proportion of sodium ions with zero water molecules in its hydration shell is negligible, but ∼10% of sodium ions have no glu bonds, reflecting the proportion of sodium ions that form 6 (fully hydrated) water bonds. The tiny fraction of sodium ions with just one water partner is consistent with the concentration and density of glu and water molecules as it would be extremely difficult to place a sodium ion such that it had only one water neighbour.

The high proportion of sodium ions bonded to two or more glu molecules will greatly limit the ions mobility, despite its small size. This may impact the overall dynamics of the system as the sodium ions may be mediating glu-glu interactions. In the next chapter the dynamics of the system as well as parallel tempering simulations are considered to address this.

3.3.4 Conclusions for the classical glutamate simulation structure

The work presented here considers the solution structure of a glu/sodium/water system. The two types of glu-glu interactions available in these systems behave differently as there are two types of carboxyl oxygen atoms available for glu-glu hydrogen bond formation. Differences between the two carboxyl groups were also present in the glu-sodium ion and glu-water interactions. The backbone (C_α) carboxyl forms fewer glu-glu and glu-water bonds than the side-chain carboxyl, but more glu-sodium bonds. Despite the different behaviour of the carboxyl groups, both formed bifurcated bonds with other glu and water molecules, with the two oxygen atoms from the same carboxyl being shared between the same positive atom. For glu-glu bonds, the C_α-carboxyl has a higher proportion of bifurcated bonds than the side-chain carboxyl, and also formed a greater proportion of the glu-sodium bonds. This suggests that either there is a greater energetic barrier to the O_C atoms forming bonds with water and glu molecules, or that the sodium ions inhibit this type of interaction and the system is sufficiently slow-moving that this was not overcome in the length of the simulation.

Another key difference between the C_α and the side-chain carboxyl is the presence of an intermediate peak in the $O_C - O_W$ RDF. A similar peak was identified by Daub *et al.* [45], which they attributed to a sodium ion mediating the oxygen-water interaction for the C_α-carboxyl. We did not find such an interaction in our simulation trajectory, but several possible motifs were identified that could have contributed to the intermediate peak. In particular, the key motif that can only be formed by the C_α-carboxyl consists of a water molecule bonding to the amine group and the C_α-carboxyl on the same glu molecule. We found that water molecules reside in this region, but are not strictly speaking 'hydrogen-bonded' to the C_α-carboxyl. Instead the water-hydrogen is orientated away from the carboxyl, stretching the $O_C - O_W$ radial distance. The stability of this motif, resulting in the well-defined shoulder in the RDF, suggests that the water molecule is stabilised by the presence of the O_C atom. It may be attempting to orientate itself into a hydrogen bond, but fails due to the presence of other atoms around the carboxyl.

We have also considered interactions between the sodium ions and the surrounding glu and water. Sodium ions rarely exist in a fully hydrated state

with a full complement of six water neighbours. The dominant feature involves sodium ions bonding to four water molecules as well as one or two glu molecules. Sodium ions never exist with zero water neighbours, suggesting that at this concentration having sodium ions completely surrounded by glu molecules is energetically unfavourable.

As previous studies of aqueous based sodium ion simulations have identified that sodium ions are usually hydrated by 6 or more water molecules, finding that the most common hydration in this system is just four waters highlights the impact of including glutamate in such systems. As mentioned by Boström *et al.* studying sodium ions in systems similar to the biological environment may significantly effect the behaviour of the sodium ions [50]. Glutamate transport in neurones and glia in the brain is thought to be driven by the cotransport of Na^+ ions [133]. Rapid information transfer between neurones and receptors occurs when glutamate is released from the neurones. In this context, glutamate is in fact the most abundant of all neurotransmitters, and exists as a'free' amino acid, and is essential to fast, efficient transmission. However, in brain hypoxia, massive releases of glutamate can cause severe brain damage. So any work on reducing the effects of brain hypoxia are require knowledge of how to terminate glutamate release.

Glutamate release is controlled only by removing its availability, rather than through the use of enzymes. It is removed from extracellular space by 'uptake carriers' that specifically exist to accumulate glutamate inside cells using ion gradients that are set up by a Na^+/K^+ pump across cell membranes. If this pump fails, or the ions run out, then the result is that glutamate is pushed into the extracellular space rather than removed from it, as is the case during hypoxia [133]. As our results clearly show that the presence of sodium ions clearly effects the hydration of sodium ions, compared to a simple ion/water solution, this may provide information on the mechanisms by which the sodium ions act in transport. The method by which sodium ions, not already attached to glutamate, attach and act as transporters is not clear, nor is the role of the sodium transport of glutamate prior to the formation of synapses. As this data also shows that the majority of glutamate molecules in this solution bond to at least one sodium ion, the ion pump used may rely solely on the preference of glutamate to attach itself to the ions. If the method of attaching the sodium ion to glutamate could

be elucidated it could provide an avenue of research leading to treatment for the glutamate accumulation in the extracellular space in the brain, as occurs in brain hypoxia. It would also be interesting to understanding how long the glutamate-sodium bond exists; whether or not the partners are exchanged often at this level of concentration and if the contact is too short for the membrane transfer to take place with just a single sodium ion-glutamate pairing, or if the bond is very long-lived and the sodium ions are therefore unavailable for future ion transfer as they remain attached to a single glutamate molecule over the time-scales of neurotransmission.

Our simulations suggest that the system undergoes very slow dynamics, which will presumably affect all similar MD simulations of this system. We found that the choice of starting configuration greatly affects the clustering distribution, but not the main structural features of the glu-glu and glu-water interactions. There is also stability across all three water potentials used, indicating that the structure is not affected too greatly by the choice of potential. The slow dynamics suggest an investigation should be made into the dynamical behaviour and also a parallel tempering study to elucidate different structural motifs. The lifetime of the sodium-ion bond, and the diffusion of the glutamate molecules through the solution may also provide significant information in the mechanism involved in neurotransmission. These are considered in the following chapter.

3.4 Car Parrinello simulations of aqueous glutamate solutions

Car-Parrinello *ab initio* MD (CPAIMD, or CPMD) simulations are used in conjunction with the classical simulation. The classical simulations are limited by their inherent approximations, but quantum mechanical studies, such as the CPMD simulation considered here, are limited by current computational capabilities, restricting the number of atoms in the system and the total simulation time. To aid the analysis of the CPAIMD results, they are shown along-side results from the classical simulations.

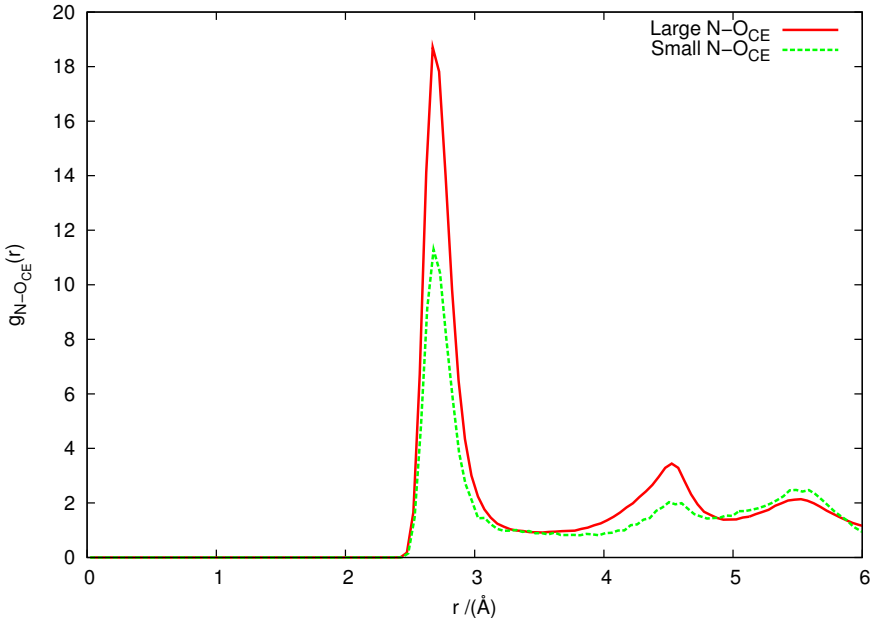
3.4.1 Structure Factor determination and validation of simulation

To establish the validity of the simulation the effects of using a small system size are considered along with comparison of the structure factor of the system with experimental results.

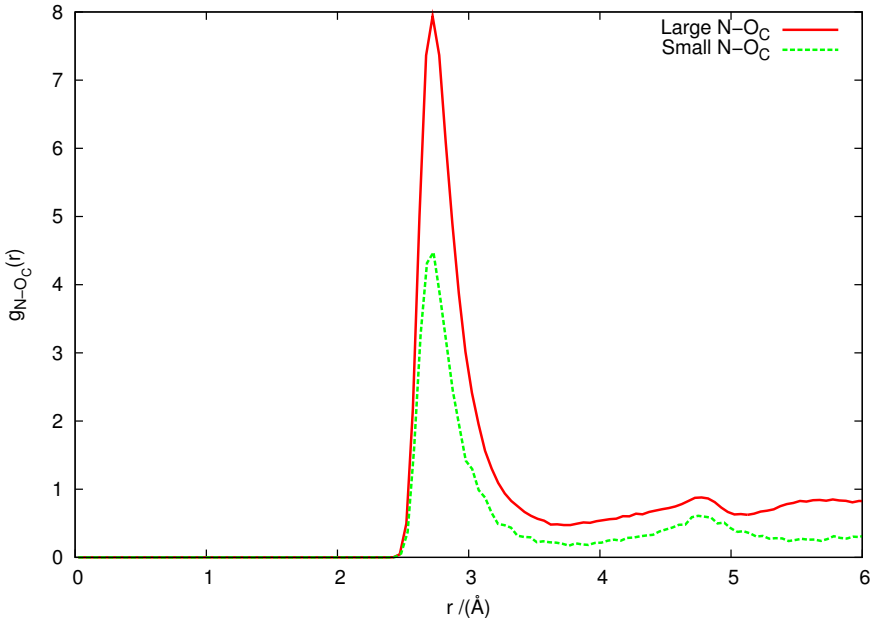
As this system is very small due to the computational limitations of performing first-principles MD calculations, finite-size effects should be considered. To establish if the small system size still produces data that provides a realistic simulation of the glu system at a concentration of 1:1:29 the radial distribution functions of two classical simulations are considered. The first is at the original size, studied classically in the previous section of this chapter. The second system is a classical box of glu molecules in solution with sodium ions and water in a smaller box of just 12.3 Å (the same size as the CPAIMD simulation that is discussed below). Both simulations were used same potentials (CHARMM22 and SPC/E) and thermodynamic conditions. These two classical simulations were chosen for comparison as the only difference between them is the system size, not the method of calculation.

Figure 3.21 shows the RDFs for the glu-glu interactions. The peaks for the small-system RDFs are much smaller than their large-system counterparts for both the $N - O_C$ and $N - O_{CE}$ interactions. This is to be expected as there are only 2 glu molecules in the small box limiting the number of glu-glu interactions available for observation. All the key peaks in the larger classical simulation (red line) are reproduced in the small system RDFs (green line) for both the $N - O_C$ and $N - O_{CE}$ interactions, suggesting that, although the number of interactions is limited, the simulation still reproduces the key behaviour of a larger system.

Figures 3.22 show the RDFs for the sodium-glu interactions for the small and large classical simulations. For both systems the sodium ion interacts with the glu molecules *via* both carboxyls, resulting in a clearly defined first peak in both plots for both system sizes. The closest agreement between the small and large box-RDFs is for the $O_{CE} - Na^+$ interaction (figure 3.22(b)), where the peak intensity is closely matched. The intensity of the first peak in the $O_{CE} - Na^+$ RDF is much smaller in the small system than the large system. As suggested by the classical results discussed at the beginning of this chapter, the system seems to be fairly slow moving. It is reasonable to suggest this difference in intensity

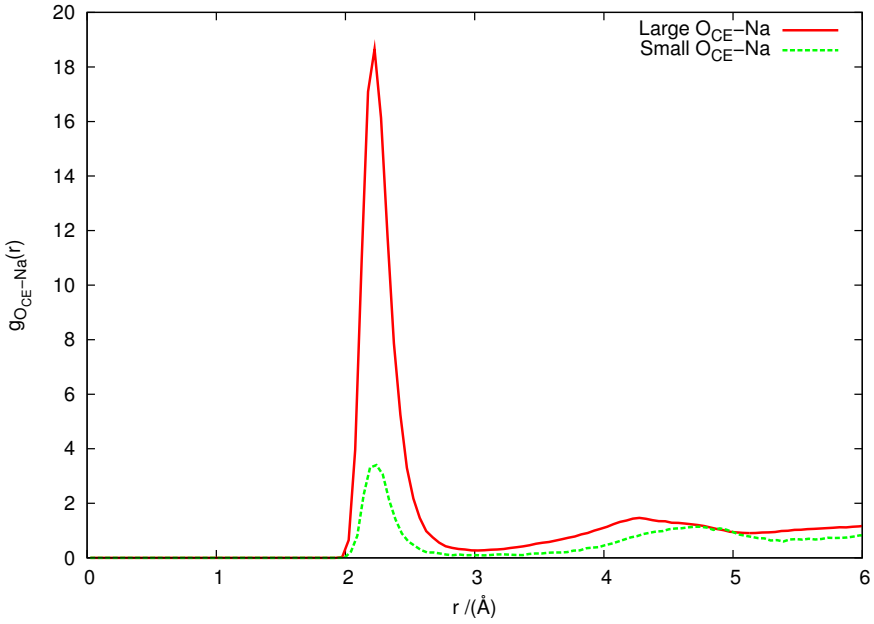


(a) $g_{N-O_{CE}}(r)$

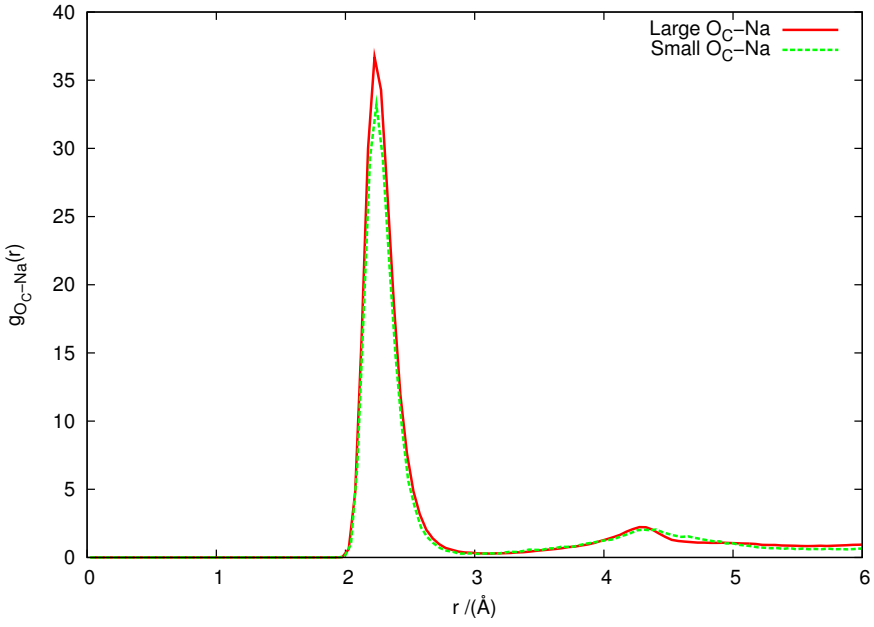


(b) $g_{N-O_C}(r)$

Figure 3.21: Glu-glu RDFs: 3.21(a) $N - O_{CE}$ RDF and 3.21(b) $N - O_C$ RDF for classical (CHARMM glu and sodium and F3C water) glutamate systems simulated in a small and large box.



(a) $g_{O_{CE}-Na^+}(r)$



(b) $g_{O_C-N_A}(r)$

Figure 3.22: Glu-sodium RDFs: 3.22(a) $O_{CE} - Na^+$ and 3.22(b) $O_C - Na^+$ RDFs for classical (CHARMM glu and sodium and F3C water) glutamate systems simulated in a small and large box.

is due to the small number of sodium ions and glu-molecules in the small system coupled with the slow dynamics inhibiting the ability O_{CE} and Na^+ atoms to interact.

The RDFs for the amine-water, $N - O_W$, carboxyl-water, $O_{CE}/O_C - O_W$, sodium-water, $Na^+ - O_W$ and water-water, $O_W - O_W$ interactions are shown in figures 3.23, figure 3.24 and 3.25. In each plot the position of the first peak coincides for the small and large systems, indicating that the small system size does not affect the ability to form inter-molecular bonds.

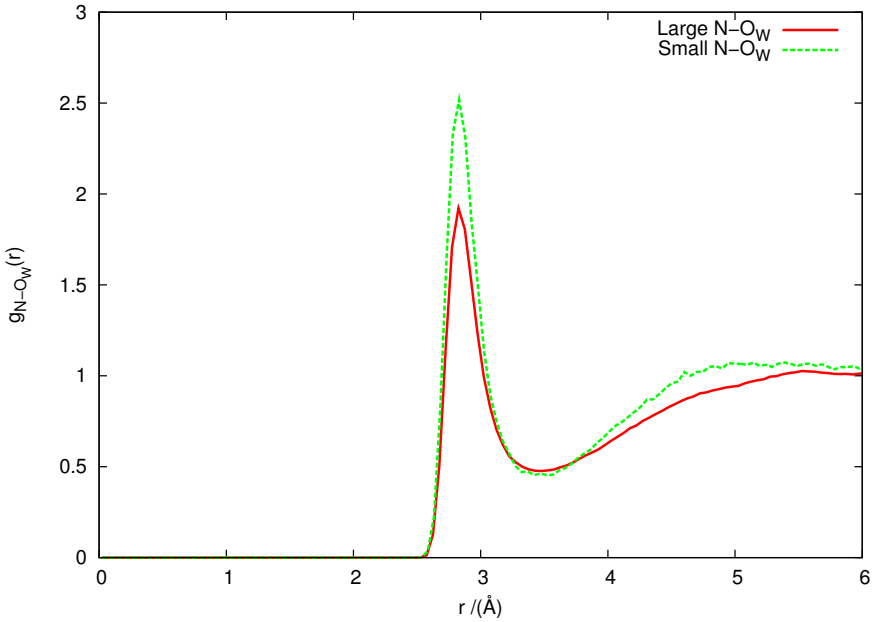
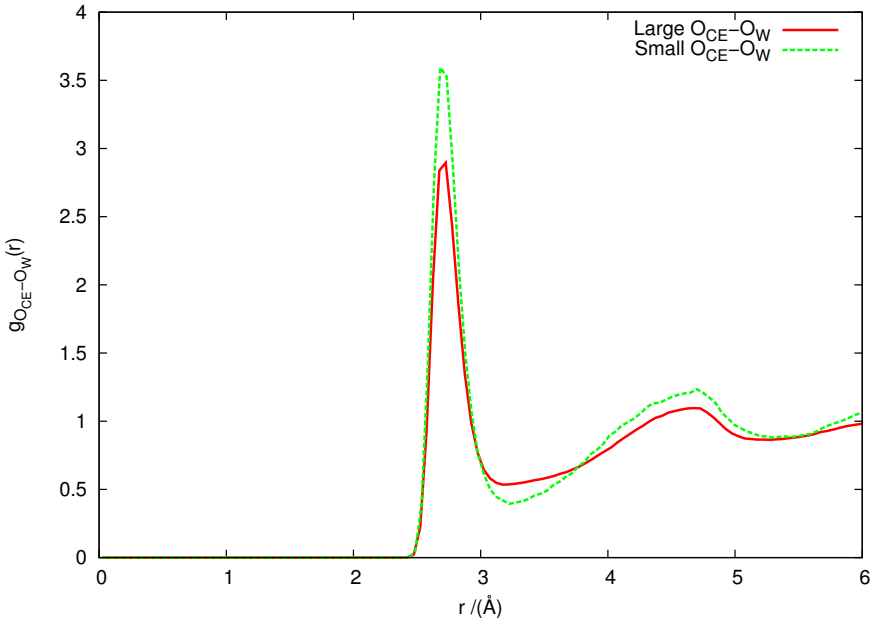


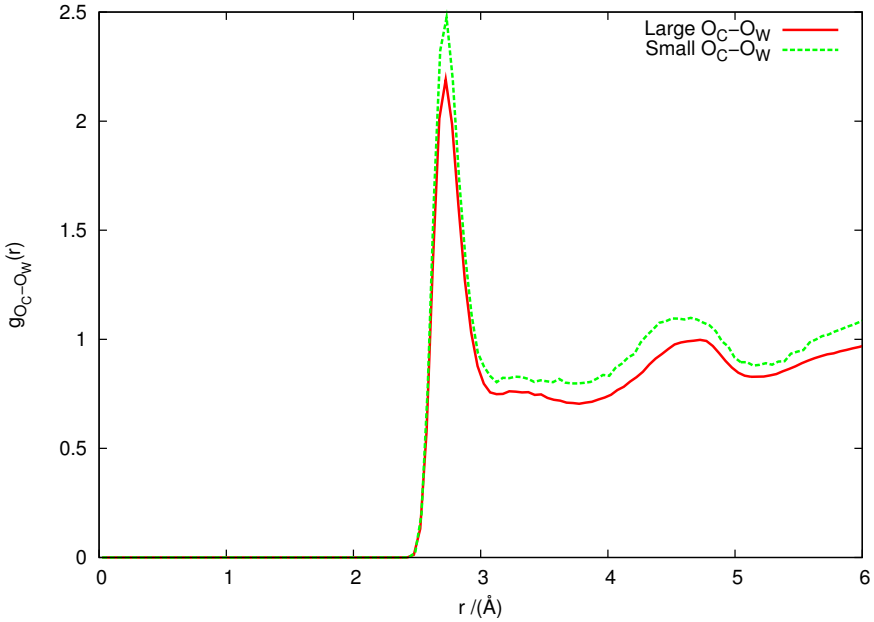
Figure 3.23: Water-glu RDFs comparing the $N-O_W$ interactions in the small classical box and the large classical box of glutamate in solution.

The water-water ($O_W - O_W$) RDF, figure 3.25(b), is almost identical in the two systems, suggesting that the small system size does not affect the bulk-water structure. The other RDFs all show a greater intensity on the first and second peaks of the RDFs in the case of the smaller system. As the glu-glu and glu-sodium RDFs all under-estimate the peaks in the small system, presumably because of the limited number of interactions of this type available, this may enable more glu-water and sodium-water interactions to occur, consequently resulting in an over-estimation of these RDF-peaks.

These classical RDFs show that despite the small system size, the key structural features identified in the classical study are still reproduced. The

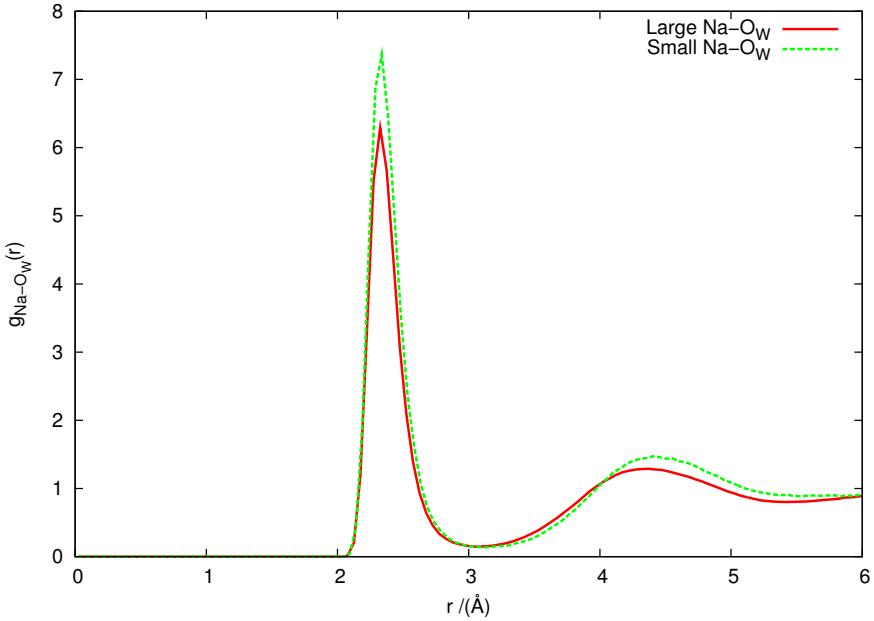


(a) $g_{O_{CE}-O_W}(r)$

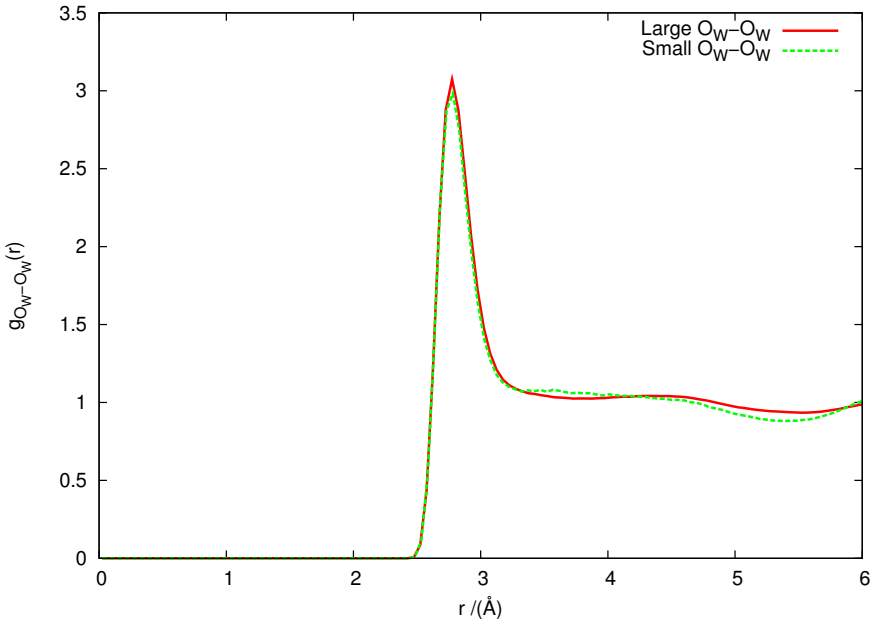


(b) $g_{O_C-O_W}$

Figure 3.24: Water-glu RDFs comparing the $O_{CE} - O_W$ and $O_C - O_W$ interactions in the small classical box and the large classical box of glutamate in solution.



(a) $g_{Na^+-O_W}(r)$



(b) $g_{O_W-O_W}(r)$

Figure 3.25: Water-sodium (Na^+-O_W) and water-water (O_W-O_W) RDFs comparing interactions in the small classical box and the large classical box of glutamate in solution.

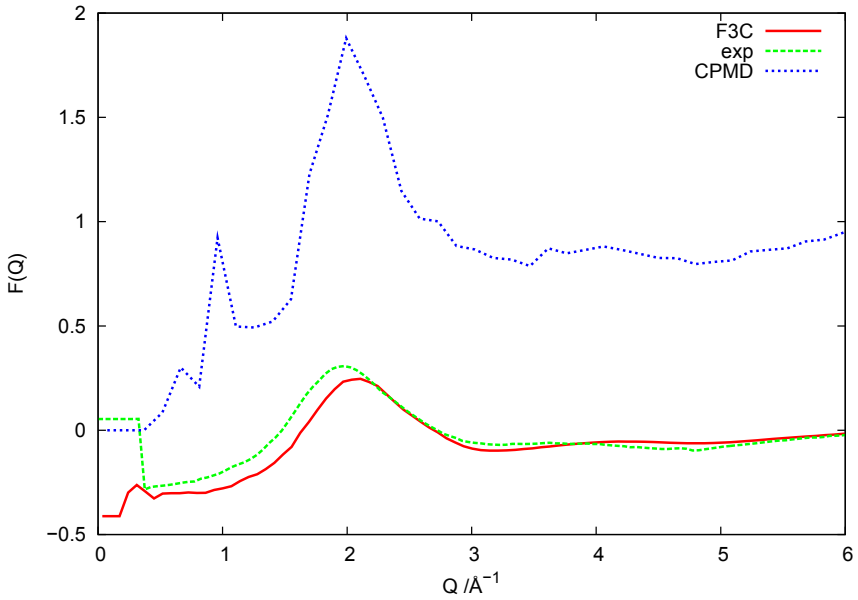


Figure 3.26: Comparison of the glu structure factors in the classical CHARMM22-F3C simulation and the CPMD system. The CPMD data has been offset by +0.5 for clarity.

size of the CPMD system is suitable, provided that the differences in glu-glu and glu-water interactions mentioned above are considered.

Further validation of the CPMD simulation is obtained by comparison of the structure factor with experimental results. The structure factor for the CPMD system is shown with the experimental $F(Q)$ obtained by McLain *et al.* [3] in figure 3.3, along with the structure factor from the classical F3C solution studied in the first part of this chapter.

The key structural features of the structure factor have been reproduced by the CPMD simulation, however there is significant over-estimation of the peaks This divergence was also seen by Daub *et al.* in their *ab initio* simulations, and they attributed it to poor sampling of the possible glu configurations, as suggested by Daub *et al.* [45]. Although the structure is over-estimate, the presence of the peaks, coupled with the limitations of the size of the simulation suggest that for the purposes of enhancing the information provided by the classical simulation, the differences found between the structure factors does not invalidate the results.

Checks on for finite-size effects find that, despite the small system size, key structures are reproduced, although there is an impact on the quality of the statistics obtained, due to the small number of interactions available. We

therefore conclude that this first-principles calculation is capable of providing an adequate description of the system.

3.4.2 Glutamate-glutamate interactions

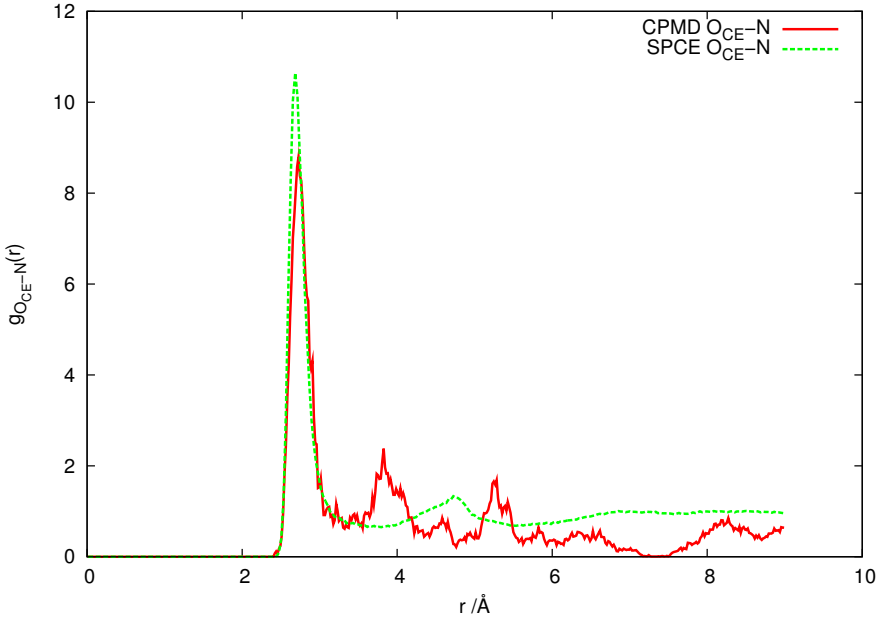
This system has just two glu molecules and therefore, as discussed above, the reported glu-glu interactions are limited in number, providing only limited value as a measure of the behaviour of glu-glu interactions. In addition, as discussed with the classical simulations, the system appears to exhibit slow dynamics. CPMD simulations are computationally expensive, limiting both the possible simulation length and size, further reducing the ability of the glu-glu interactions to provide a representative picture of a system of this size. Hence, they are considered here only as additional information to the classical simulation.

The lack of interaction between a the C_α-carboxyl and the amine for the entire 23 ps trajectory further emphasises the slow dynamics of the system and the resulting dependence of results on the initial configuration.

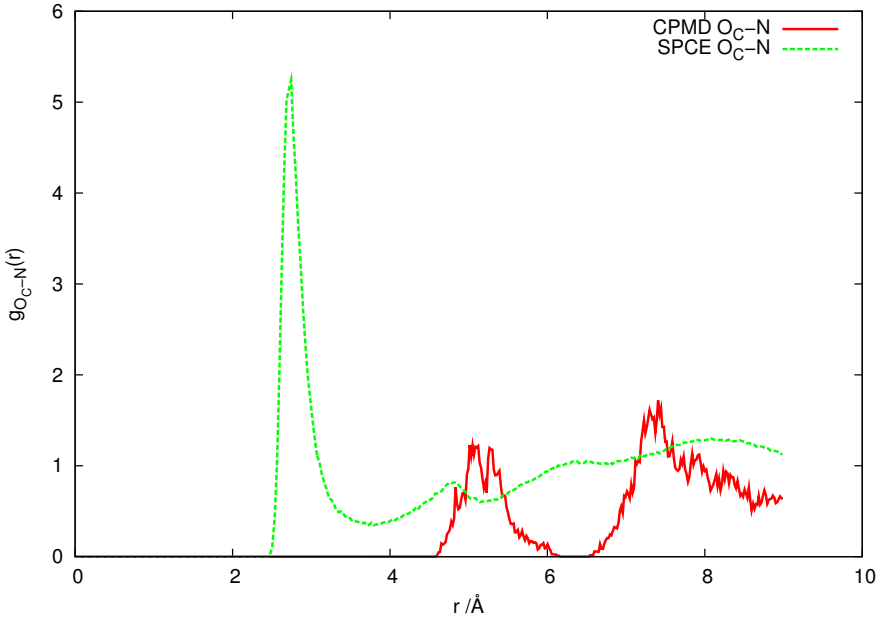
Figures 3.27 and 3.28 show the RDFs for glu-glu hydrogen bond contacts between the amine nitrogen and each of the two species of carboxyl oxygen. We find that the only glu-glu contact occurs between the side-chain (O_{CE} oxygen) and the amine, with zero intensity in the first-peak region for $g_{O_C-N}(r)$ and $g_{O_C-H_X}(r)$, indicating no direct $O_C - H_X - N$ contacts.

As glu-glu bonds occur *via* the side-chain, and in the classical study bifurcated motifs were found between the side-chain carboxyl and the amine, we next consider the 2D radial-radial distribution correlating the $r_{H_X-O_{CE1}}$ and $r_{H_X-O_{CE2}}$ radial distance in figure 3.29. For comparison the classical SPC/E-glu results are also shown in figure 3.29(a).

It is clear from the 2D radial-radial plots, is that there are no bifurcated motifs between the O_{CE} and H_X atoms, indicated by the zero intensity in the $r_{H_X-O_{CE1}} \simeq r_{H_X-O_{CE2}} \simeq 2.1$ Å. Given the length of time that these two sites are correlated, even though this is a short simulation, if bifurcated motifs were going to occur in the CPMD simulation, then we would expect to see evidence for them in this plot. The preference for short, single linear bonds in the CPMD system, where the second carboxyl oxygen co-ordinates at 3 Å rather than in a range between 2 and 3.5 Å, is very strong; the CPMD system has a highest intensity peak in the linear bond region five times greater than the peaks in the SPC/E

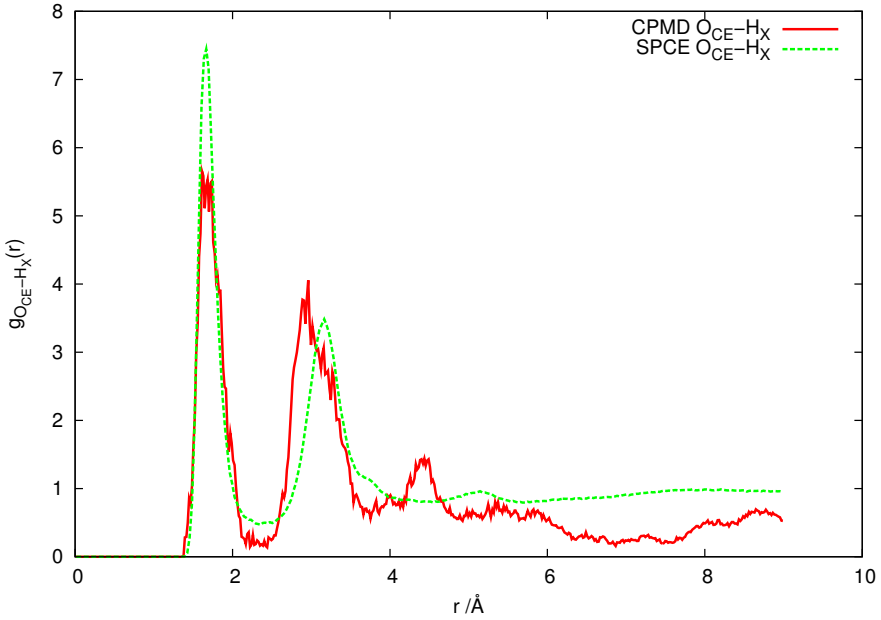


(a) $g_{O_{CE}-N}(r)$

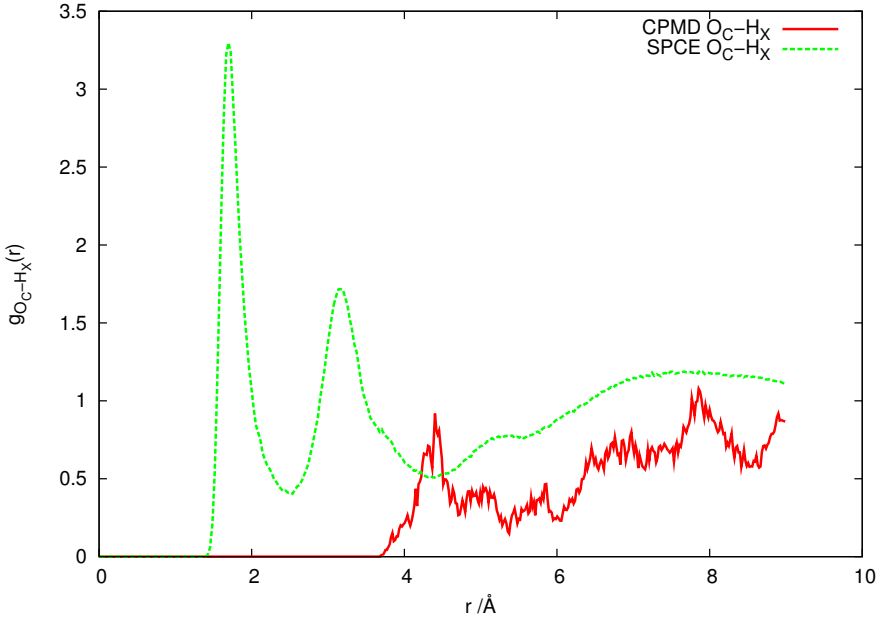


(b) $g_{O_C-N}(r)$

Figure 3.27: 3.27(a) $O_{CE}-N$ RDFs and 3.27(b) O_C-N RDFs for the CPMD system and and the classical CHARMM22-SPC/E simulation.

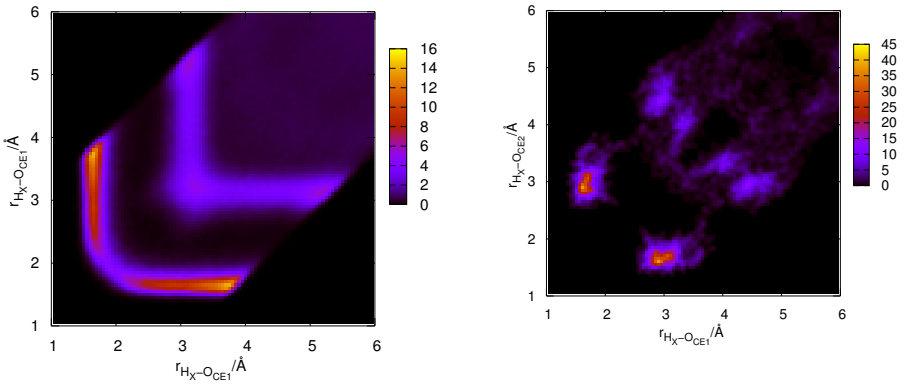


(a) $g_{O_{CE}-H_X}(r)$



(b) $g_{O_C-H_X}(r)$

Figure 3.28: 3.28(a) $O_{CE} - H_X$ RDFs and 3.28(b) $O_C - H_X$ RDFs for the CPMD system and the classical CHARMM22-SPC/E simulation.



(a) SPCE $H_X \cdots O_{CE}$ (b) CPMD $H_X \cdots O_{CE}$

Figure 3.29: 2D radial-radial distributions correlating $r_{H_X-O_{CE1}}$ and $r_{H_X-O_{CE2}}$ for the CPMD system and the SPC/E classical system.

plot. The SPC/E system has bonds throughout the single and bifurcated region. This difference between the CPAIMD and classical results can be attributed to the broader spread of possible configurations available to the classical molecule. It should be noted that the CHARMM22 potential model is known to produce more bifurcated motifs than other classical potentials [114]. In the classical study the C_α-carboxyl had a higher propensity to form bifurcated motifs than the side-chain, and therefore this does not rule out bifurcation *via* the C_α group.

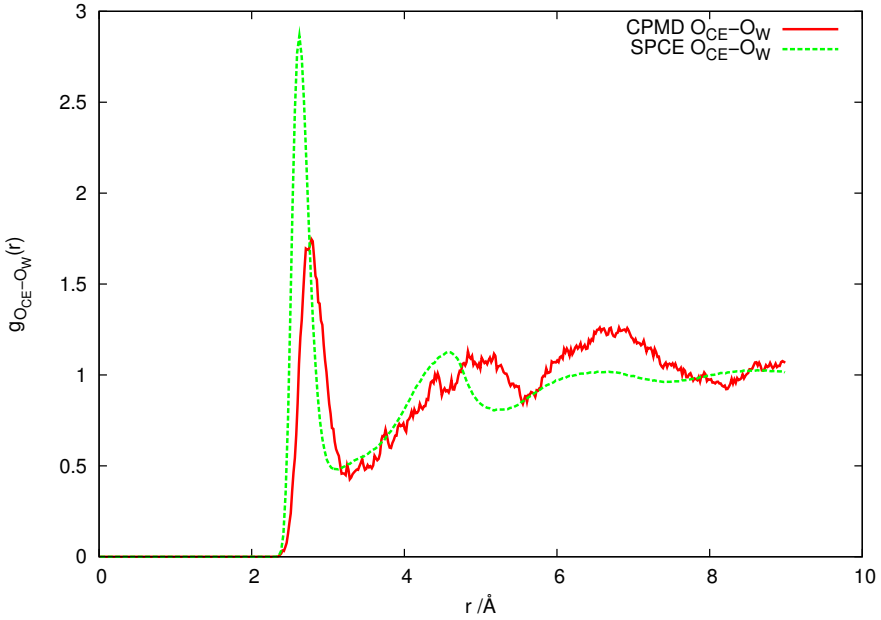
Due to the lack of statistically significant data it is inappropriate to draw too many conclusions from the glu-glu interaction data presented here. Thus, we will not persue the analysis of the glu-glu interactions any further. The CPMD data does emphasise the importance of the initial configuration, and verifies the structure of the O_{CE}-amine RDFs obtained from the classical simulations. As CPMD is a 'first-principles' simulation method, although not perfect, these results aid the validation of the classical simulation.

3.4.3 Water structure in glutamate solutions

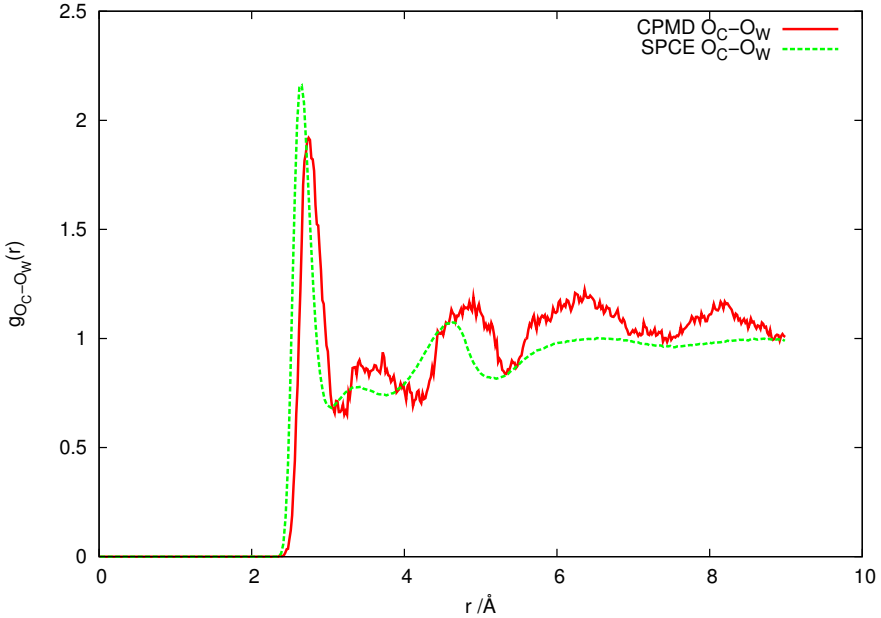
The water-glu interactions again suffer from poor statistics, although less significantly than the glu-glu interactions. There are enough water molecules to fully hydrate all of the charged groups of each glu molecule, enabling a good understanding of the interactions. Figure 3.30 shows the interaction between

the two types of carboxyl oxygen, the side chain (O_{CE}) and the C_α (O_C), with the water-oxygen (O_W). Both plots show that the O_{CE}/O_C-water hydrogen bond, indicated by the first peak, is present in the CPMD solution as well as the SPC/E system. The plots do not show the marked difference in intensity that the classical study described, although the first-peak intensity is smaller in the $O_{CE} - O_W$ interaction than the $O_C - O_W$ interaction. The $O_{CE} - O_W$ and $O_C - O_W$ co-ordination numbers are 3.12 and 2.92 respectively, showing a slight increase in water-carboxyl bonds around the side-chain. In comparison, the classical results provided co-orindation number of 2.62 and 2.19 respectively, again showing a decrease, but of compared with the classical results of 16% compared to 6% in the CPMD results. Daub *et al.* also found little difference in intensity between the side-chain and C_α distribution functions for their first-principles simulation results, despite observing the difference in the classical data [45]. Daub *et al.* note that the peaks of their first-principles simulation results are significantly 'shallower and broader' than their classical counterparts. We do not find significantly broader peaks. The only key difference is the shift of the glu-water RDF peaks to slightly larger r in the CPMD results. This is most likely an artefact of the Lennard-Jones potentials used in the classical parameterisation.

The key difference between the two carboxyl-oxygen to water-oxygen RDFs arise in the region $r > 3$Å. In the classical study the C_α-carboxyl RDF has a shoulder between the first and second peaks, which is not present in the side-chain RDF. In the CPMD study this difference is further emphasised as the shoulder appears as a peak in its own right in the $O_C - O_W$ RDF, while being completely absent form the $O_{CE} - O_W$ RDF. We also see a slight shift of all the peaks to larger r in the CPMD study, suggesting slightly longer hydrogen-bond interactions. This constasts with the results obtained by Leenders *et al.* [44] where the bond lengths shifted to smaller r in their first-principles simulations, although they only considered one glu molecule in 135 water molecules, and thus the difference between the results reported here and those obtained by Leenders *et al.* could be an artefact of system concentration. It should be noted that the shift of the peaks depends on the classical simulation that it is compared with, and is therefore not directly comparable between these different studies. Daub *et al.* [45] also found little co-ordination outside the initial hydration shell in their first-principles results for either carboxyl group.



(a) $g_{O_{CE}-O_W}(r)$



(b) $g_{O_C-O_W}(r)$

Figure 3.30: 3.30(a) $O_{CE} - O_W$ RDFs and 3.30(b) $O_C - O_W$ RDFs for the CPMD system and the classical CHARMM22-SPC/E simulation.

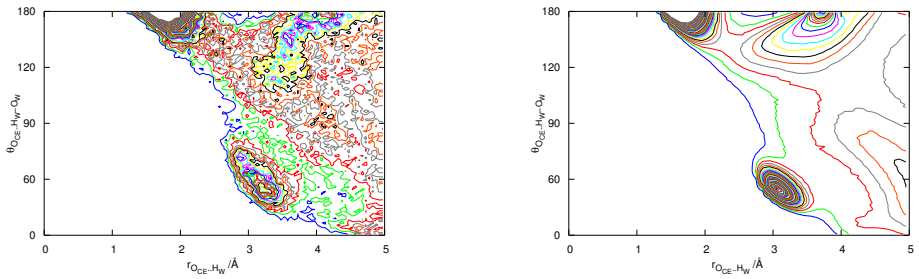
In the classical study, the shoulder between the first and second peaks of the $O_C - O_W$ RDF was attributed to the presence of a 'circular' structure formed by a water molecule bridging the gap between the C_α-carboxyl and the amine within the same glu molecule. On analysis of the trajectory file, using the same hydrogen bond cut-off distances as the classical study, we found that of ∼6.6 million $O_W - O_C$ and $\sim 73,000$ $O_W - N$ hydrogen bonds, there were 27,906 circular motifs (0.4% of the $O_C - O_W$ bonds and 38% of the $O_W - N$ bonds), using the same cut-off distances as for the classical simulation.

The classical results showed that 0.4% of the $O_C - O_W$ bonds and 8.4% of the $O_W - N$ bonds formed the circle motif. As the CPMD results shift the shoulder to larger r, we also consider extending the cut-off distances to encompass the entire 2nd peak in the CPMD $O_C - O_W$ RDF ($3.78 < r < 4.00$ Å) and to allow for the broader first peak of the CPMD $N - O_W$ RDF ($3.48 < r < 3.60$ Å). This increases the percentage of circle motifs to 0.5% of the CPMD $O_C - O_W$ bonds and 46% of $N - O_W$ bonds, as shown in table 3.5, indicating that only a slightly greater proportion of circles are included in this range. We also checked for the case suggested by Daub *et al.* where a sodium ion could be mediating the water-glu interaction. As with the classical study, we found no evidence of this motif in the CPMD system.

	No. of Circles	No. $N - O_W$ bonds	% circles to $N - O_W$	No. $O_C - O_W$ bonds	% circles to $O_C - O_W$
Classical	55,008	657,525	8.4%	12,530,985	0.4%
CPMD Short cut-offs	27,906	73,307	38%	6,621,626	0.4%
CPMD Long cut-offs	36,675	78,202	46%	6,673,640	0.5%

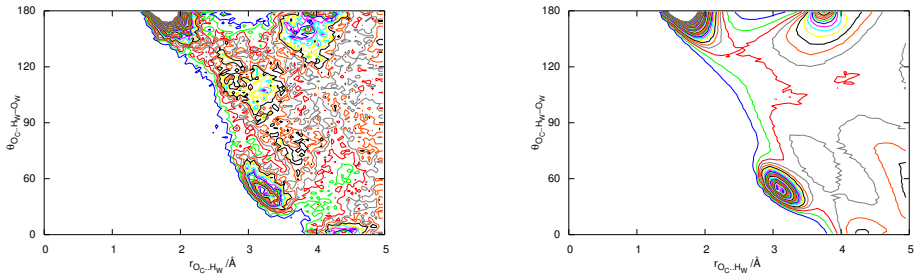
Table 3.5: Number of $O_C - O_W$ and $N - O_W$ hydrogen bonds formed and the number of circle motifs that these resulted in, for the classical simulation (ref [4]) and the CPMD simulation, with the cut-offs set to the same as the classical and then extended to the outer edge of the second peak of the $O_C - O_W$ RDF and the broader first peak of the $N - O_W$ RDF.

To further our understanding of the different behaviour of the two carboxylate groups we consider the 2-dimensional radial-angular and radial-radial distribution functions for the water-carboxyl interactions. Figure 3.31 shows the classical and



(a) CPMD $O_{CE} \cdot\cdot H_W - O_W$ (b) SPC/E $O_{CE} \cdot\cdot H_W - O_W$

Figure 3.31: 2D radial-angular distribution plots for the CPMD solution 3.31(a) and the classical SPC/E solution 3.31(b), correlating the distribution of $O_{CE} - H_W - O_W$ angles to the $O_{CE} - O_W$ radial length.

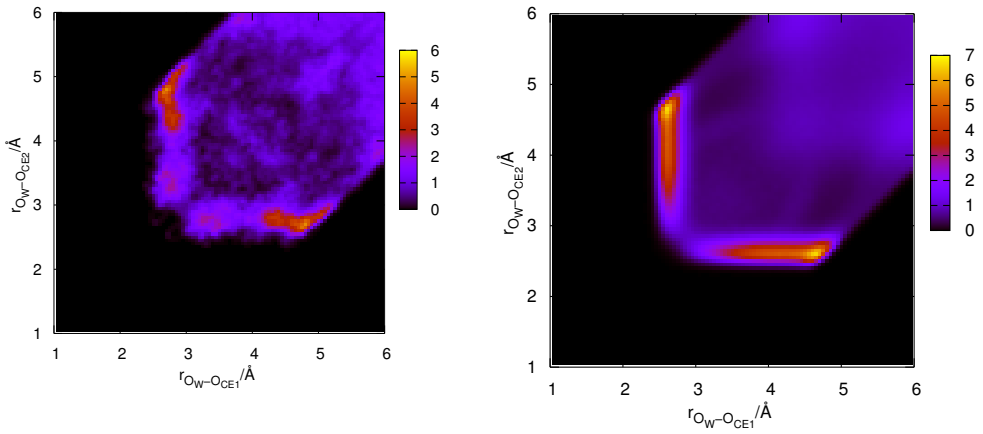


(a) CPMD $O_C \cdot\cdot H_W - O_W$ (b) SPC/E $O_C \cdot\cdot H_W - O_W$

Figure 3.32: 2D radial-angular distribution plots for the CPMD solution 3.32(a) and the classical SPC/E solution 3.32(b), correlating the distribution of $O_C - H_W - O_W$ angles to the $O_C - O_W$ radial length.

CPMD 2D-radial-angular distributions comparing the radial $O_{CE} - O_W$ distance with the $O_{CE} - \widehat{H_W} - O_W$ angle. Figure 3.32 shows the corresponding plots for the $O_C - O_W$ interaction. It should be noted that the apparent difference in drawing style between the classical and CPMD plots is due to the poor sampling of the CPMD system resulting in data that is not 'smoothed', in a similar fashion to the less smooth curves of the CPMD RDFs above.

Both the carboxyl groups show similar 2D radial-angular distributions, with a peak at $r \simeq 1.7$ Å, $\theta \simeq 180°$ corresponding to a short linear hydrogen bond, as identified in the classical plots. The second pronounced feature at $r \simeq 3.0$ Å, $\theta \simeq 60°$ is due to the co-ordination of the second carboxyl oxygen with the water

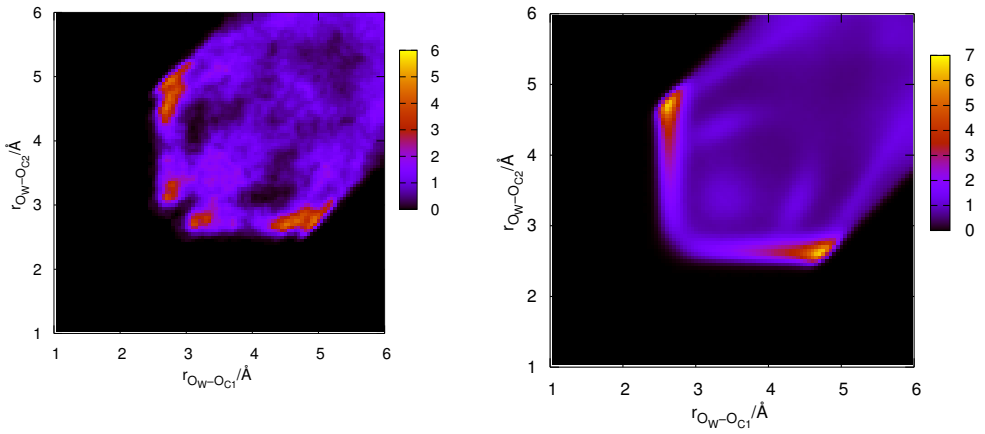


(a) CPMD $O_W \cdots O_{CE}$ (b) SPC/E $O_W \cdots O_{CE}$

Figure 3.33: 2D radial-radial distribution plots for the CPMD solution 3.33(a) and the classical SPC/E solution 3.33(b), correlating $r_{O_W-O_{CE1}}$ and $r_{O_W-O_{CE2}}$.

molecule bonded to the first carboxyl oxygen within the same molecule. Again this feature is present for both carboxyls and for both the classical and CPMD simulations, validating the classical results. In both the classical and CPMD plots we see broader peaks in the O_{CE} distributions than the O_C distributions, with the effect being most pronounced in the CPMD system. Broadening of peaks in glu-glu interactions in the classical solutions can be attributed to the greater proportion of bifurcated bonds enabling a broader range of stable motifs. To assess if the same is true in this instance we consider the 2D radial-radial distributions correlating the $r_{O_W-O_{CE1}}$ and $r_{O_W-O_{CE2}}$ distances in figure 3.33, and the $r_{O_W-O_{C1}}$ and $r_{O_W-O_{C2}}$ distances in figure 3.34.

For both the $O_W - O_{CE}$ and $O_W - O_C$ 2D radial-radial plots (figures 3.33(a) and 3.34(a) respectively), we find regions of intensity in the CPMD plots that replicate those in the classical SPC/E plots, indicating similar bonding motifs. There is a complete absence of intensity in the true bifurcated region at $r_1 = r_2 = 2.8$ Å in the CPMD plots, which is present in the classical plots. The area immediately surrounding this is also considered a bifurcated region and it is not zero intensity, suggesting that some motifs exist where a water molecule is simultaneously bonded to both carboxyl-oxygens. As seen in the classical case, the majority of bonds are single, linear hydrogen bonds, with the second carboxyl oxygen coordinating at a distance of 4 Å or greater, as highlighted by



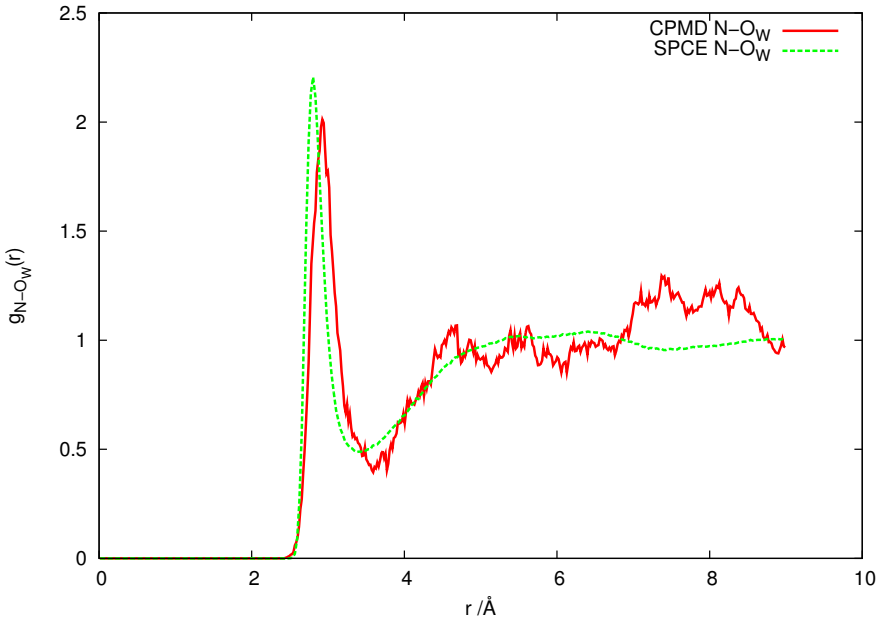
(a) CPMD $O_W \cdot\cdot O_C$ (b) SPC/E $O_W \cdot\cdot O_C$

Figure 3.34: 2D radial-radial distribution plots for the CPMD solution 3.34(a) and the classical SPC/E solution 3.34(b), correlating $r_{O_W-O_{C1}}$ and $r_{O_W-O_{C2}}$.

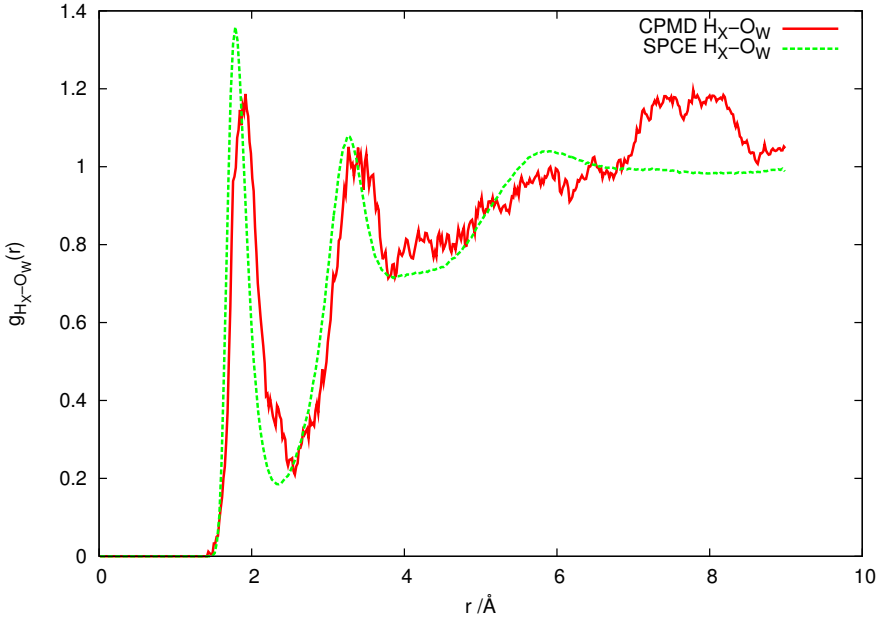
the area of greatest intensity in the regions $r_{O_W-O_{C1}} \simeq 3$ Å,$r_{O_W-O_{C2}} \simeq 4.8$ Å, and $r_{O_W-O_{C1}} \simeq 4.8$ Å, $r_{O_W-O_{C2}} \simeq 3$ Å (and similarly for O_{CE} atoms).

Figure 3.35 plots the RDFs for the amine-water interaction ($N - O_W$ and $H_X - O_W$). In the CPMD system, both the $N - O_W$ RDF (figure 3.35(a)) and the $H_X - O_W$ RDF (figure 3.35(a)) interaction show peaks indicating the presence of a $N - H_X \cdot\cdot O_W$ hydrogen bond, that correspond to the peaks in the SPC/E RDFs. As seen in the other RDFs, there is a shift to larger r. There is also no clear second hydration shell, as shown by the classical simulations. As with the carboxyl-oxygen-water interactions the co-ordination number for $N - O_W$ increases from 2.59 in the classical F3C system to 3.83, indicating more glu-water interactions. This is to be expected as there is a decrease in glu-glu interactions due to the small system size, and therefore a corresponding increase in the number of glu-water interactions.

Figure 3.36 shows the water-water distribution functions, comparing the CPMD results with the corresponding RDFs from the classical SPC/E system. We find good agreement in the placement of the peaks for both the $O_W - O_W$ and $O_W - H_W$ interactions. As with the glu-water interactions there is an increase in the water-water $O_W - O_W$ co-ordination from 6.07 in the F3C classical system to 7.91 in the CPMD simulation. As this occurs for glu-water and water-water interactions, and there are very few glu-glu interactions, this may mean that

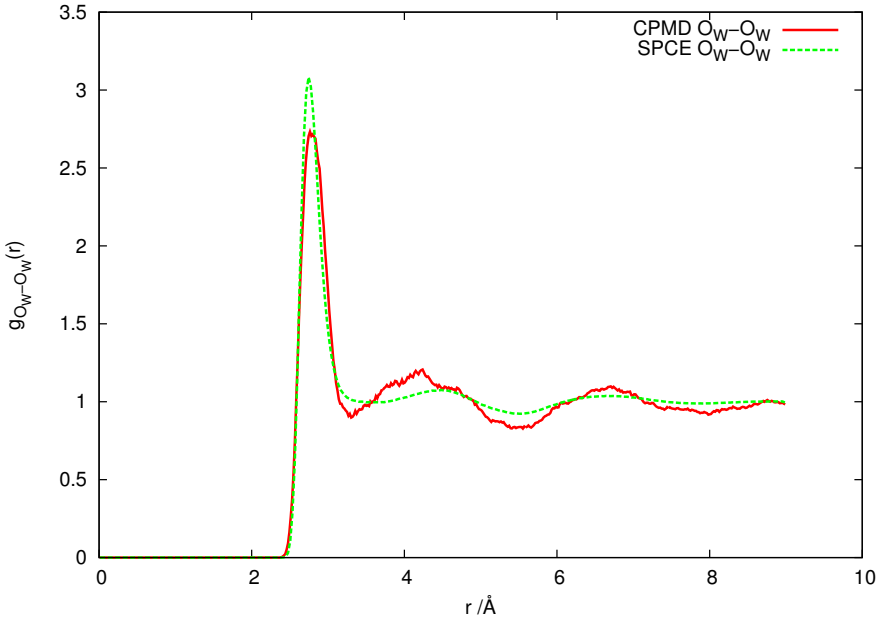


(a) $g_{O_W-N}(r)$

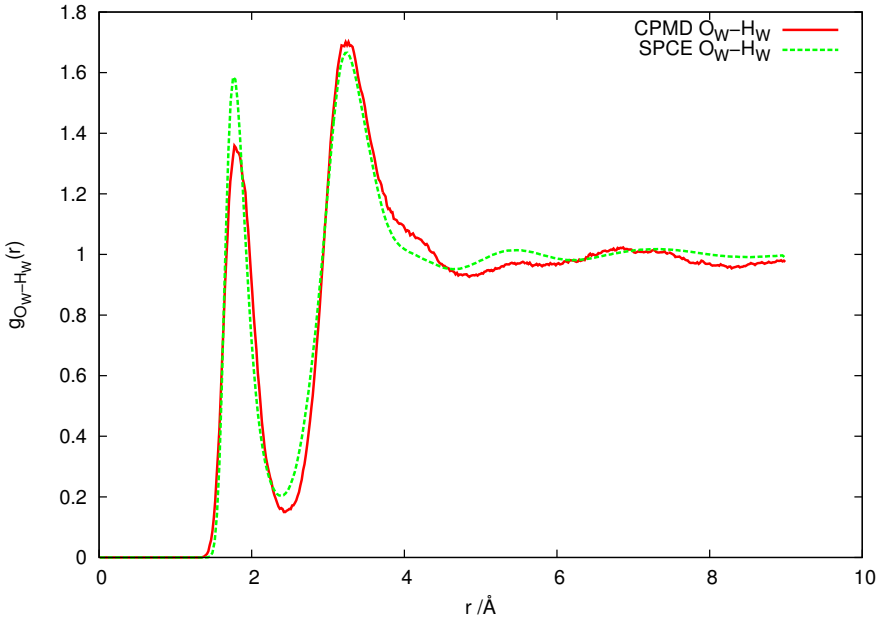


(b) $g_{O_W-H_X}(r)$

Figure 3.35: 3.35(a) $O_W - N$ RDFs and 3.35(b) $O_W - H_X$ RDFs for the CPMD system and the classical CHARMM22-F3C simulation.

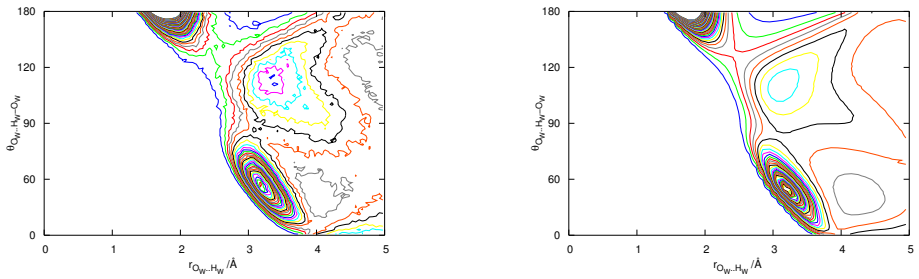


(a) $g_{O_W-O_W}(r)$



(b) $g_{O_W-H_W}(r)$

Figure 3.36: 3.36(a) $O_W - O_W$ RDFs and 3.36(b) $O_W - H_W$ RDFs for the CPMD system and the classical CHARMM22-F3C simulation.



(a) CPMD $O_W \cdots H_W - O_W$ (b) SPCE $O_W \cdots H_W - O_W$

Figure 3.37: 2D radial-angular distribution plot from the CPMD (figure 3.37(a)) and classical SPC/E (figure 3.37(b)) systems, correlating the $O_W - H_W - O_W$ angle to the $O_W - O_W$ radial length.

the CHARMM22 force-field under-structures the system compared to the CPMD approach. It should be noted that over-structuring of water has been suggested as an artificial artefact of first principles calculations of water [134], and not necessarily representative of the true system nature. However, as there is also a difference in the glu-water structure, although to a lesser degree, this may not be the only explanation for this change.

Figure 3.37 shows the 2D radial-angular distribution plots correlating the $O_W - O_W$ radial distance and the $O_W - \widehat{H_W - O_W}$ angle. Again, the key structural features of the classical plots are reproduced in the CPMD plot, with peaks at $r \simeq 1.7\mathring{A}$, $\theta \simeq 180°$ (the strong hydrogen bond) and $r \simeq 3.1\mathring{A}$, $\theta \simeq 60°$ (the interaction of the second unbonded hydrogen).

Unlike the carboxyl interactions, there is no broadening of the peaks in the water-water CPMD 2D radial-angular distribution plots compared with the classical plots. This fits in with the good re-production of the RDFs indicating only minor changes in the water structure, indicating that there is little difference in the 'strength' of the bonds as there is still the same freedom of movement. The sharpening of the RDFs, and the increased water structure that this would suggest should be shown by a tightening of the peaks in the 2D radial-angular distribution. We conclude that any difference between the classical and CPMD water structures is minimal.

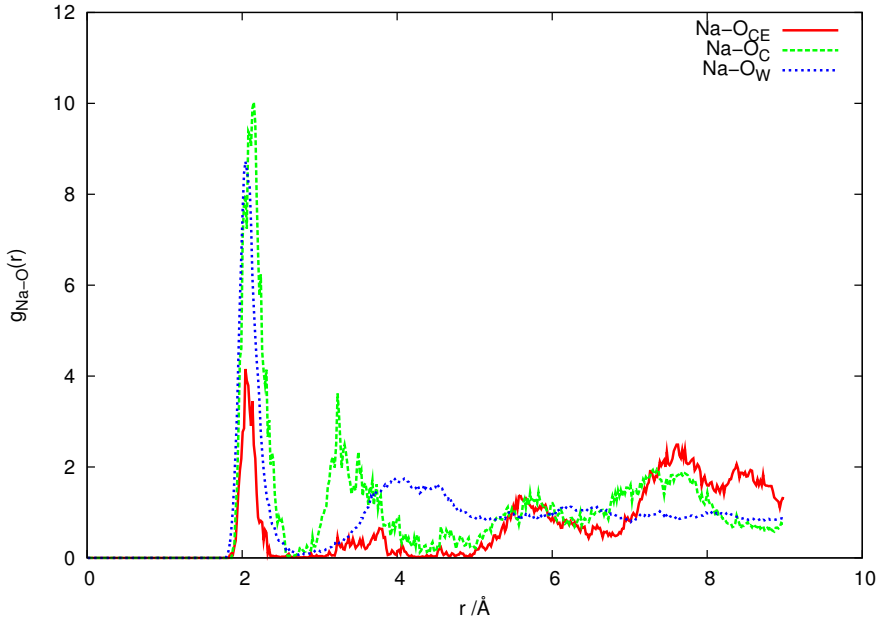
3.4.4 Sodium interactions with glutamate and water

Figure 3.38 shows the RDFs for sodium ions interacting with each of the carboxyl oxygens, O_C and O_{CE}, and water-oxygen, O_W, for the classical SPC/E system and the CPMD system. Both simulations show that the dominant sodium ion interaction is with the C_α-carboxyl oxygens, O_C, as indicated by the greatest intensity in the first peak. The CPMD simulation show a change in the proportion of $O_W - Na^+$ and $O_{CE} - Na^+$ bonds from those observed in the classical results, with far more $O_W - Na^+$ interactions in the CPMD simulation. The results are of poor statistical merit as there are only two sodium ions and two glu molecules, and therefore this change may be due to the inability for the glu molecules and sodium ions to approach *via* the side-chain, thereby increasing the relative proportion of water-sodium ion interactions. The sodium ions and glu molecules may take a long time to diffuse through the system, reducing their interaction with other bonding sites during the short simulation time available.

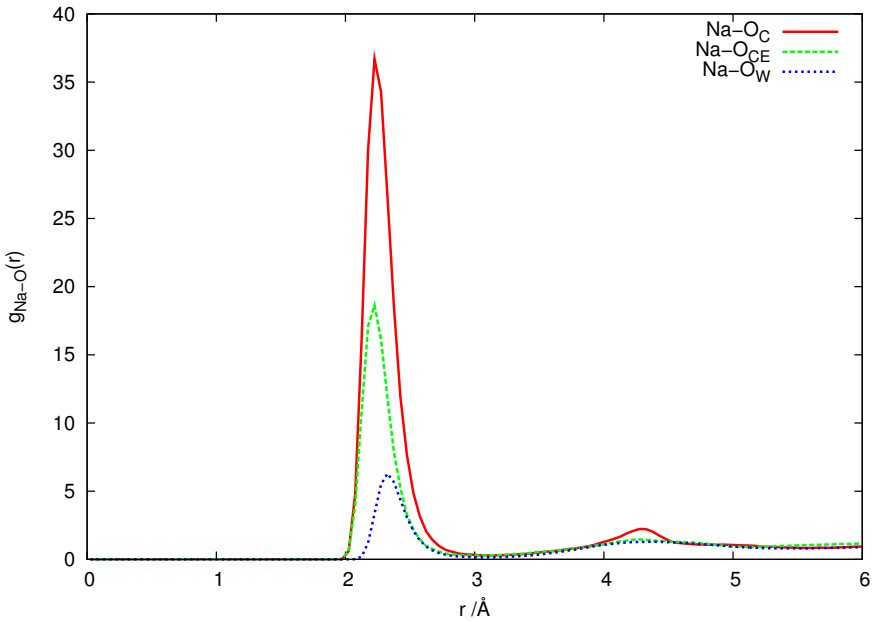
One key difference between the classical and CPMD results is the clear development of a second hydration shell at $\sim$ 4 Å that is not present in the original simulation, indicating that the sodium ions perturb the water far more in the CPMD simulations. As mentioned before, this could be an artefact of the over-structuring of water that can occur in first principles simulations [134].

Interestingly, despite the shift of peaks to larger r in the other RDFs, the sodium ion RDFs in figure 3.38 have the first peak shifted to smaller r for all three interactions plotted. This indicates that the potential exerted by the sodium ion on the surrounding molecules is significantly different from the classical case. As the classical results depend on an empirical force-field to characterise these interactions, this suggests that, in the instance of the glu system, the sodium ions are not correctly parameterised with respect to the other components of the system, although it does not appear to effect the overall behaviour of the system. As the other RDFs shift the peak to larger r, it is not clear whether the glu, water or sodium ion potentials are incorrect. The choice of potential used for the sodium ion is discussed more in the following chapter and in appendix D.

Finally, we show the hydration of sodium ions by water and also the number of glu-molecules surrounding each sodium ion in figure 3.39. These numbers are the percentage of sodium ions (averaged over all ions in the system and over all sampled time-steps of the simulation) with 0, 1, ...6 bonded neighbours (either of

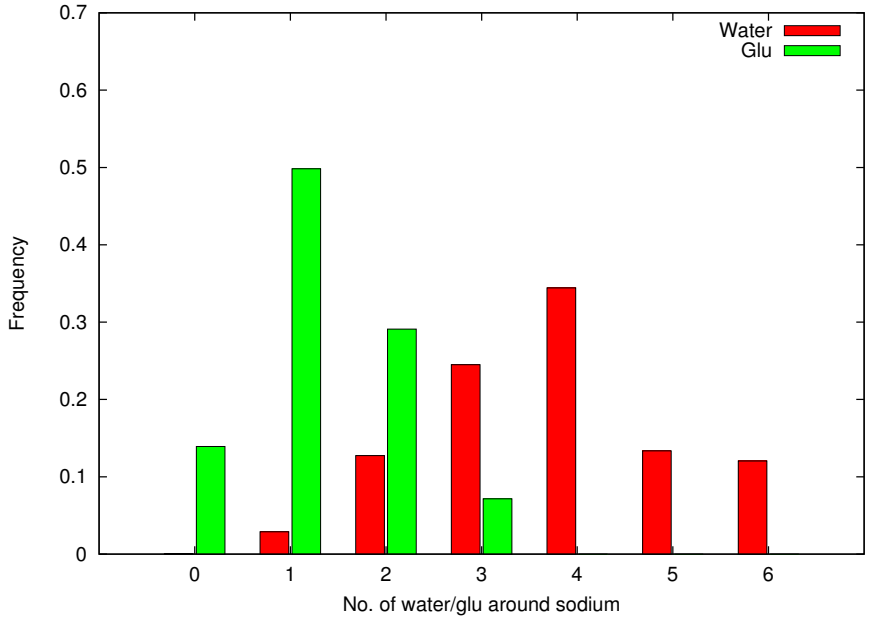


(a) CPMD

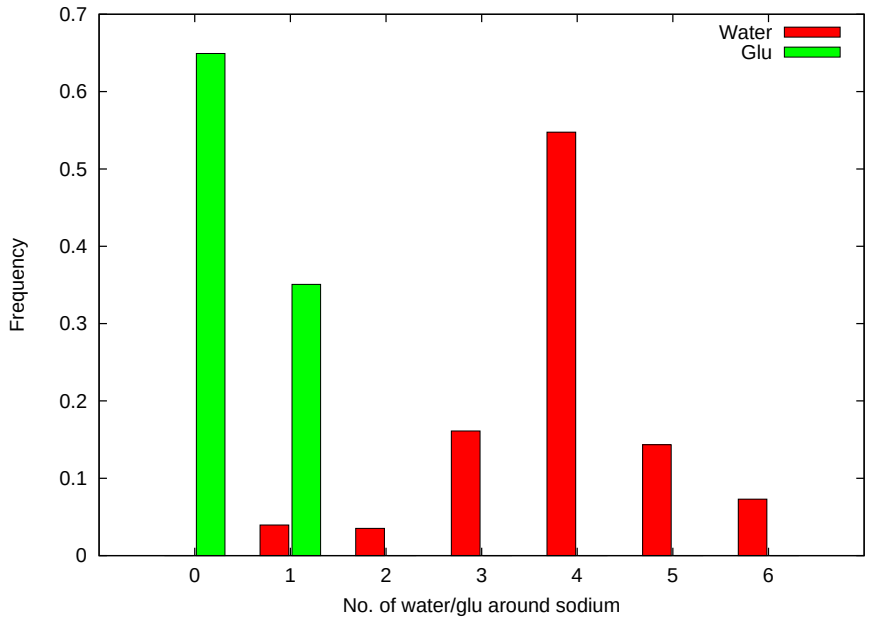


(b) SPC/E

Figure 3.38: RDFs for the interaction of sodium ions with O_C, O_{CE} and O_W for the CPMD (3.38(a) and for the SPC/E classical simulation 3.38(b).



(a) Classical



(b) CPMD

Figure 3.39: Hydration (as a percentage of total number of ion bonds of type water-sodium or glu-sodium) of sodium ions by water and glutamate molecules for the classical simulation 3.39(a) from ref [4] and for the CPMD simulation 3.39(b).

type O_W or of type O_C/O_{CE}). The term 'neighbour' is used here to refer to an atom in close-enough proximity that it sits within a radial distance less than or equal to the trough of the first minimum in the corresponding RDF.

The CPMD system sodium ions have significantly fewer glu-neighbours, and consequently a higher proportion of water neighbours compared to the classical results. No sodium ions have more than one glu neighbour. The maximum that the system could achieve would be two neighbours. We might therefore expect that more sodium ions are 'fully' hydrated with 6 water molecules (sodium ions can fully hydrate with a coordination number of 6.53 according to a classical study on aqueous NaCl [51]). There is actually a decrease in the percentage of fully hydrated sodiums ions from 12% in the classical simulation to 7% in the CPMD simulation, and a corresponding decrease in the proportion of sodium ions with 4-water neighbours. The lower sodium-glu occupation numbers and the increase in proportion of sodium ions with four neighbours may explain the increased structure in the second and third hydration shells of sodium ions.

3.4.5 Summary conclusions of CPMD simulations

The CPMD simulation aimed to validate and verify some of the structural features extracted from the classical study at the beginning of this chapter. We were able to validate the simulation, although the limited system size and simulation length resulted in poor statistics. The structure factor in both the CPMD and classical simulations under-estimates the low-Q peaks, indicating that poor sampling of the configuration space is most likely the reason for the low-peaks, rather than the choice of classical potential.

Although we considered the glu-glu structures only briefly, due to there being two-molecules in the system, it highlighted that the system appears to exhibit slow dynamics. We were able to ascertain that the $O_{CE} - H_X$ interaction did not result in bifurcated bonds, despite their prevalence in the classical study. In contrast water-carboxyl interactions indicated the presence of bifurcated bonds for both carboxyl species, indicating that in this instance their presence in the classical study was not an artefact of the CHARMM22 potential. We also found a greater prevalence of the 'circle'-motif identified in the classical study, with the 'shoulder' in the $O_C - O_W$ RDF becoming a clearly distinct peak in the CPMD results.

In all but the sodium ion RDFs, there was a shift to larger r from the classical study. However, in the case of any RDF involving the sodium ion, the shift was to smaller r, indicating a constriction of bond lengths. This could be an artefact of the sodium ion potential used in the classical study, although it is not clear at this time. The shift had no obvious effect on the overall physical structure of the system.

The hydration of sodium ions are of particular interest in this system, as although the concentration is the same as the classical system, the small system size could vary the sodium-glutamate interactions. At no time in the entire simulation did both glutamate molecules form bond with the same sodium ion. This could be due to the short simulation time, or that this is an unfavourable configuration. It is unlikely that it is entirely due to it being unfavourable as the classical simulation provided instances of sodium ions being surrounded by 2 and 3 glutamate molecules. We also see a greater proportion of sodium ions preferring four water neighbours than in the classical simulation, with the majority of the change occurring by reducing the proportion of sodium ions with only two or three water molecules in its hydration shell. As smaller numbers of water molecules in the hydration shell necessarily requires a larger number of glutamate molecules in the hydration shell this is not surprising given the that sodium ions only interact with one glutamate molecule at a time at most in the CPMD simulations.

As the hydration of sodium ions by glutamate, limited to just one neighbour, is different from the classical simulations, this may suggest either a flaw in the classical force-field parameterisation, or finite-size effects that may impact in a sightly bound biological system such as the extracellular space where glutamate molecules accumulate to act as neurotransmitters [133], or that the simulation time is too short. In each instance, this warrants further investigation, but highlights the importance of not just system concentration as discussed by Bostöm *et al.* [50], but also possibly system size. As the extracellular space is limited, this may impact on glutamate transport and therefore neurotransmission. This is work that has so far not been considered by any of the other studies on aqueous glutamate solutions.

These results once again highlight the dependence of the simulation on the initial configuration used, indicating a susceptibility to slow molecular movement. In the next chapter we further examine the dynamics of the system and perform

parallel tempering to enhance the conformational sampling.

Chapter 4

Beyond standard MD simulation: application to aqueous glutamate

The structural study of aqueous glutamate (glu) solutions in chapter 3 indicated that the results were heavily dependent on the initial configuration of the system and, therefore, exhibited slow dynamical behaviour, in particular, slow diffusion throughout the system. The time-scales available to classical molecular dynamics simulations result in relatively poor sampling of the possible conformations of the glu system. This is a problem inherent in classical simulations of many complex systems, especially bio-molecular systems where the energetic barriers between conformations may be sufficiently large that they inhibit the system moving from one configuration to another.

To further the understanding of the aqueous glu solutions the dynamics of the system are considered in the first part of this chapter, to formally assess the apparent slow-movement of the system. In the second part of this chapter the enhanced sampling technique, parallel tempering, is used to extract further configurational information on the classical simulation of glutamate (glu).

4.1 Dynamics of aqueous glutamate

In the following section the hydrogen-bond lifetimes and the self-diffusion coefficients of molecules in the classical glu simulations are considered. The analysis is performed on the set of simulations obtained for the structural study discussed in chapter 3; the three glu systems, CHARMM22 glu with sodium ions

and F3C, SPC/E or TIP3P water are considered (referred to hereafter as Glu-F3C, Glu-SPCE and Glu-TIP3P) and the three pure-water systems (F3C, SPCE and TIP3P).

Hydrogen-bond lifetimes have been used extensively in the study of aqueous molecular dynamics as a measure of the freedom of movement of a system. A variety of different approaches are used in discerning the hydrogen-bond (relaxation) lifetime. Both time-dependent and time-independent auto-correlation functions are used, i.e. those that consider the history between time 0 and time t and those that just consider times 0 and t. In addition, different methods are used to describe hydrogen bonds, either geometric or energetic, and within these criteria differing cut-offs are used, even for similar systems (for example see refs [119, 128, 135]). Some studies consider multiple geometric criterion, others a simple single radial cut-off. Further discrepancy is also introduced by the empirical potentials used as these may effect the dynamics as much as the molecules considered. The use of different time-steps, both in terms of the accuracy of the simulation, and also the sampling time-step (the frequency at which the trajectory is printed to a file for post-simulation analysis) will both effect the hydrogen-bond lifetimes. We are not currently able to compare dynamical results obtained from molecular dynamics simulations with experimental data on dynamics as none has been published on the glutamate system at this time. The aim of the following discussion is to assess the relative dynamical freedom of the glu systems and their dynamical freedom relative to pure water systems that used the same simulation parameters and empirical potentials.

4.1.1 Hydrogen-bond lifetimes

There are multiple definitions of a 'hydrogen bond' in molecular dynamics simulations. The subject of this work is not to examine this definition, particularly as it is already the subject of much debate [117, 119, 122, 123, 135]. To enable understanding of the dynamical nature of the glu systems a definition of hydrogen-bonding is required. In simulations of aqueous hydrogen-bonded liquid geometric criteria have often been adopted in previous studies (for example see refs [21, 123, 128, 135]), and this is therefore the direction taken here.

There are two common geometric definitions of a hydrogen bond, both of

which are considered by Nieto-Draghi *et al.* [128]. The first considers only the radial distance between the donor oxygen and the acceptor oxygen (for example two oxygen atoms in adjacent water molecules). The second definition extends this by also considering the radial distance between the hydrogen and the acceptor oxygen atom, as well as an angular cut-off for the $O-H\cdots O$ angle. On observation of the glu system we were able to identify situations where two atoms that have the ability to form hydrogen bonds could be in close enough proximity to fulfill the requirements of the first definition of a hydrogen bond, without the mediating hydrogen between them. This particularly occurs in water molecules around the carboxyl sites of the glu molecule. In the following analysis we choose the second, stricter definition of a hydrogen bond such that molecule A and molecule B are considered hydrogen-bonded if all three of the following criteria are met:

1. $r_{O_A-O_B} < r_{OO-cut}$
2. $r_{H_A-O_B} < r_{OH-cut}$
3. $\theta_{O_A-H_A-O_B} < \theta_{cut}$.

The radial cut-off distance is taken as the first-minimum in the relevant RDF (for example for the hydrogen bond between two water molecules the O-O, $r_{O_A-O_B}$, and H-O, $r_{H_A-O_B}$, RDFs are considered). Forrconsistency the average position of the first-minima across the three types of glu system simulated is used. The radial cut-off distances used in this discussion are given in appendix B, table B.1.

The angular cut-off is commonly set at 30° as molecules that obey the first two radial cut-offs almostalways approach a linear configuration [123]. To confirm this for the glu simulations, figure 4.1 plots the frequency of angles for bonds obeying the first two radial criteria in the Glu-F3C system. Similar distributions were obtained for the other systems. This shows that the majority of interactions occur in the region above 170°, and therefore an angular cut-off of 30°, where the interaction angle X-Y-Z must be greater than or equal to 150°, is reasonable for this study (it should be noted that the y-axis has an exponential scale for ease of plotting the data). The plot shows non-zero values below angles of 170°, although in each instance the proportion of the interactions where this is the case is small in comparison to the number above 170°.

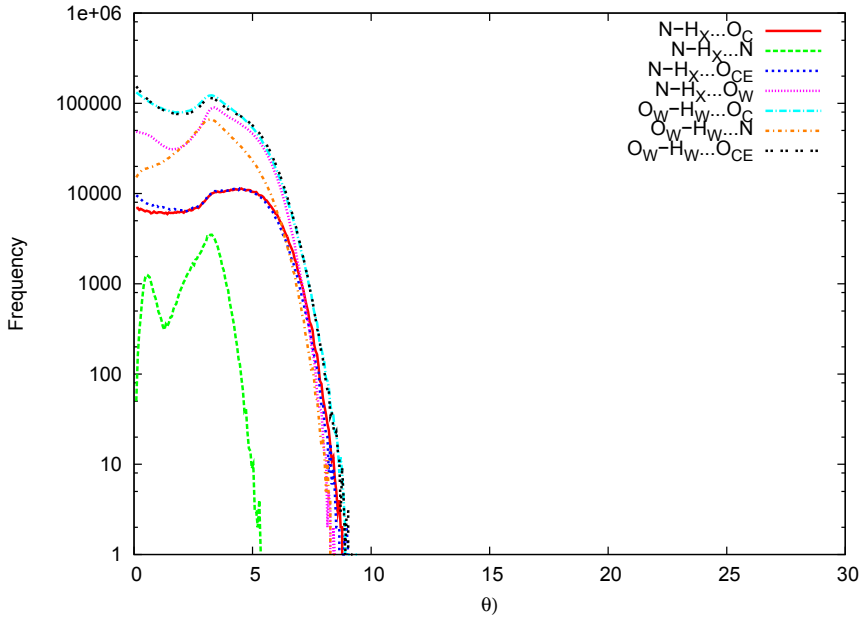


Figure 4.1: The distribution of angles of bonding interactions that satisfy the first two radial criteria.

The main limitation of these hydrogen-bond criteria is the resulting short lifetimes. For example, Nieto-Draghi *et al.* quote time-dependent hydrogen-bond lifetime in pure water as 2.2 ps using the single O-O cut-off as the bonding criteria [128]. Using this definition we obtain a lifetime in the pure SPC/E solution of 2.0 ps. The stricter, three-fold definition of the hydrogen-bond results in a hydrogen-bond lifetime of just 0.08 ps. In comparison quasi-elastic incoherent neutron scattering experiments obtain a relaxation time of 1.25 ps for pure water [136]. As this investigation is concerned with the relative behaviour of the systems considered here, this short lifetime is not considered an issue, especially as close proximity of the two oxygen atoms, alone, does not constitute a hydrogen bond.

This three-fold criteria is used for all oxgyen-hydrogen-oxygen hydrogen bonds, and also for the bonds involving the amine group of the glu molecule where O_A and H_A are instead the N and H_X atoms (refer to the labelling scheme used for the glu simulations in chapter 3). We also consider the dynamics of the sodium ions in the solution by using a simple geometric definition where the sodium ion is considered bonded to atom X if $r_{Na^+-X} < r_{Na^+-X-cut}$, where r_{Na^-X-cut} is the radial position of the first minimum of the relevant RDF.

Using the three-fold geometric definition of a hydrogen-bond, a time-dependent auto-correlation function, $C_{HB}(t, t^*)$, is constructed, as described in section 2.3.6, where t^* enables temporary breaking and re-formation of the hydrogen bond [123, 137, 138, 128]. This is then fitted to an exponential function to identify a lifetime, $\tau_{HB}(t^*)$, as described by Nieto-Draghi *et al* [128]:

$$C_{HB}(t, t^*) \sim exp(\frac{-t}{\tau_{HB}(t^*)}), \tag{4.1}$$

For comparison we have then calculated the ratio $\frac{\tau_{HB}(t^*)}{\tau_{pure}(t^*)}$, where $\tau_{pure}(t^*)$ is the water $O_W - H_W \cdots O_W$ lifetime in the pure SPC/E water system, using $t^* = 0, 0.05, 0.1$ and 0.5 ps. The hydrogen-bond lifetime for the SPC/E pure water $O_W - H_W \cdots O_W$ interaction is $\tau_{pure}(t^* = 0) = 0.1$ ps. The full set of hydrogen bond lifetimes are provided in appendix B, table B.2. We have used the SPC/E pure water system lifetimes in comparison with all three classical glu-solution systems.

Figure 4.2 plots $\frac{\tau_{HB}(t^*)}{\tau_{pure}(t^*)}$ (the relative hydrogen bond lifetime) for different values of t^* for the glu-glu interactions for each of the three systems considered. In all cases the glu-glu bond-lifetimes are greater than those of the pure-water system, as indicated by relative hydrogen-bond lifetimes greater than 1. For each interaction we see an increase in the relative bond lifetime with increasing t^*. The greatest value of t^* results in the greatest difference between the pure water hydrogen-bond lifetime and the glu-glu bond-lifetimes. This indicates that the glu-glu bonds are more easily able to re-form than water. This may be partly due to the large molecule size limiting the ability of the molecules to move apart, and therefore the bonds re-form due to their close proximity, whereas water is more able to diffuse away and completely break the hydrogen bonds. Using a value of t^* that is too large may be misleading as it may result in some of the data including bonds that really are broken and reformed being interpreted as one long-lived bond.

The two F3C system interactions, $O_{CE} - N$ and $O_C - N$, both have similar trends across the different values of t^*, as do the interactions for the SPC/E and TIP3P systems. There is closer agreement between the two types of interaction within a system rather than between one type of interaction across the three systems. This suggests that the water potential impacts the glu-glu bonding

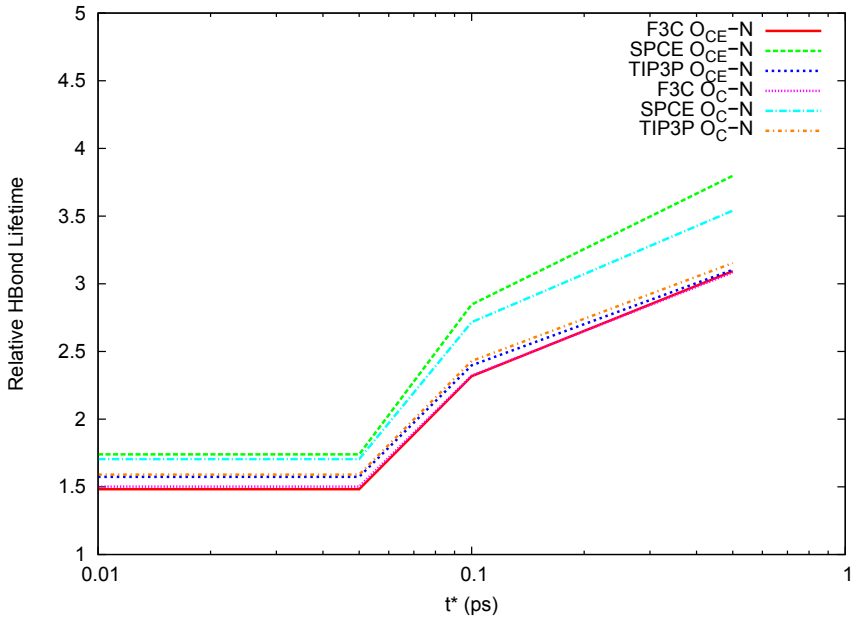


Figure 4.2: The relative hydrogen-bond lifetime, $\frac{\tau_{HB}}{\tau_{pure}}$ for the glu-glu interactions between the two carboxyl groups and the amine in the three glu systems, for $t^* = 0.01$, 0.05, 0.1 and 0.5 ps.

more than the difference between the two carboxyl sites.

Figure 4.3 plots the hydrogen bond lifetime, $\tau_{HB}(t^*)$ for the glu-glu interactions in the F3C system. As expected from the relative hydrogen-bond lifetimes in figure 4.2 we see that the $O_{CE} - N$ and $O_C - N$ have very similar lifetimes across all values of t^*, with both increasing at a similar rate with increasing t^*. This figure also plots the glu-water hydrogen bonds, $O_{CE} \cdots H_W - O - W$, $O_C \cdots H_W - O_W$ and $N - H_X \cdots O_W$, and the water-water hydrogen bond lifetimes $O_W \cdots H_W - O_W$ for the F3C-glu system. The $O_{CE} - O_W$ and $O_C - O_W$ plots are similar, again with increasing t^*. In each instance the carboxyl-water lifetimes are shorter and increase less rapidly with t^* than the glu-glu lifetimes.

The water-water lifetimes are similar to those of the $N - O_W$ (amine-water) interaction, indicating a similar strength interaction for $O_W - O_W$ and $N - O_W$, that is more easily broken than the carboxyl-water bonds. The lifetime remains similar for all values of t^* for the $N - O_W$ and $O_W - O_W$ bonds suggesting that these interactions rarely have temporary breaks in the hydrogen-bond that rapidly re-form. Similar trends were found for the SPC/E and TIP3P glu systems.

Figure 4.4 shows the hydrogen-bond lifetime for the glu-water interactions

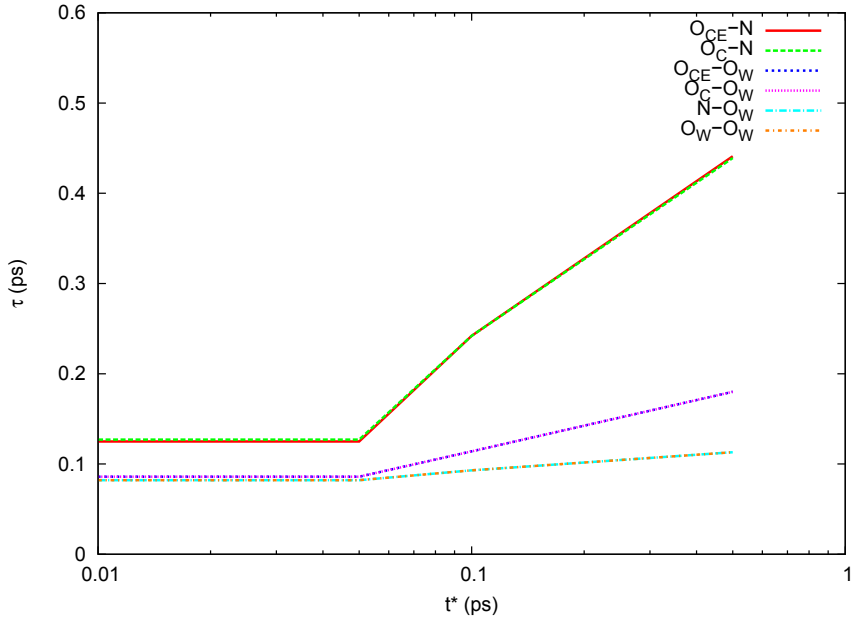


Figure 4.3: $\tau_{HB}(t^*)$ for the glu-glu, glu-water and water-water interactions in the F3C/CHARMM22 classical simulation of glu discused in chapter 3, for $t^* = 0.01$, 0.05, 0.1 and 0.5 ps.

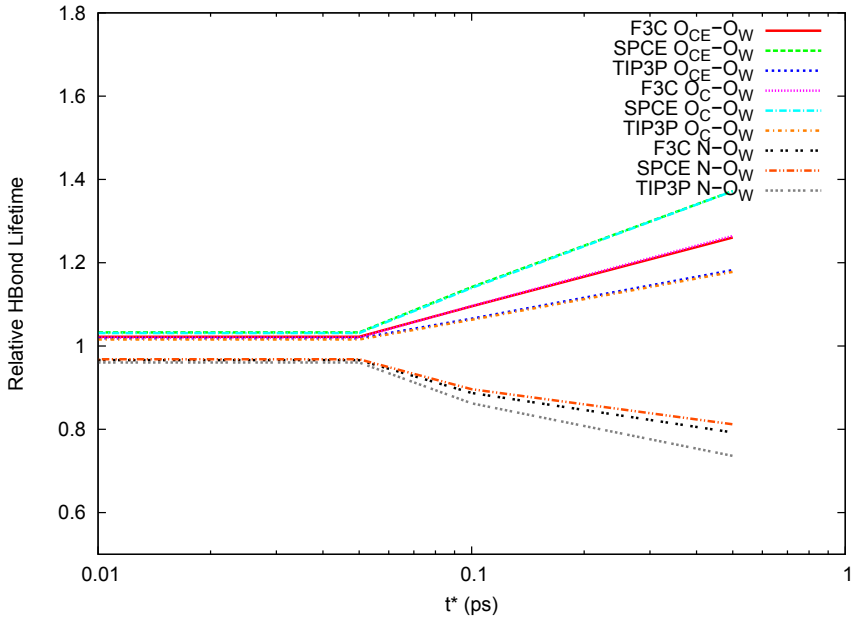


Figure 4.4: The relative hydrogen-bond lifetime, $\frac{\tau_{HB}}{\tau_{pure}}$ for the glu-water interactions in the three glu systems, for $t^* = 0.01$, 0.05, 0.1 and 0.5 ps..

relative to the hydrogen-bond lifetime of the F3C pure water system. For the carboxyl-water ($O_{CE}-O_W$ and O_C-O_W) interactions, the hydrogen bond lifetime is consistently greater across all values of t^* than the pure water system O_W-O_W lifetimes (indicated by the values greater than 1). The lifetimes also increase faster than the pure water lifetime with increasing t^*. The $N-O_W$ (amine-water) lifetime is smaller than the O_W-O_W lifetime, and decreases relative to the pure water lifetime with increasing t^*, indicating that the $N-O_W$ interaction is less susceptible to temporary breakages than pure-water. For all 6 types of interaction shown in these plots we find that the SPC/E-glu system has the greatest relative hydrogen-bond lifetime, suggesting that the SPC/E water potential reduces the mobility of the components of the system, although the effect is small.

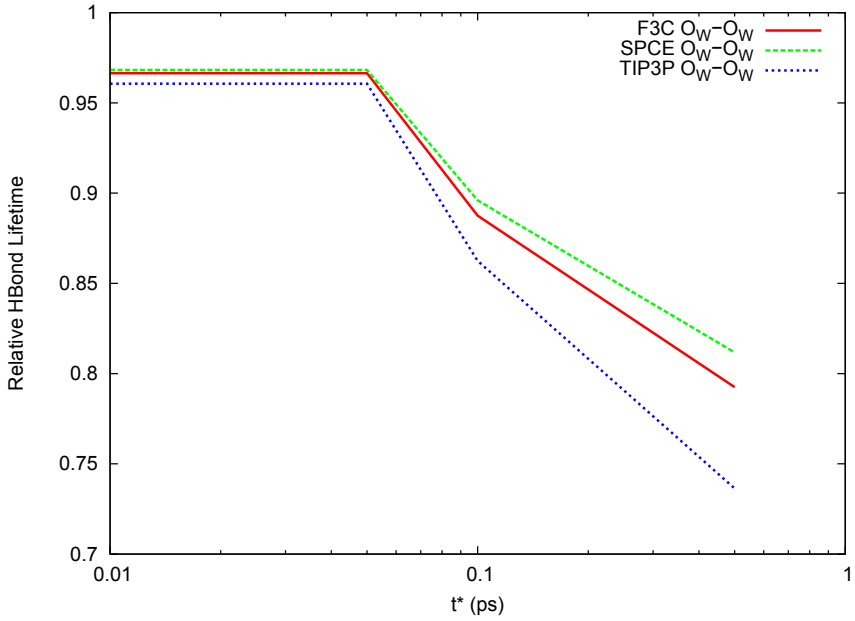


Figure 4.5: The relative hydrogen-bond lifetime, $\frac{\tau_{HB}}{\tau_{pure}}$ for the water-water hydrogen bond, $O_W - H_W \cdots OW$, in the three glu systems, relative to the F3C pure-water hydrogen bond lifetime, for $t^* = 0.01$, 0.05, 0.1 and 0.5 ps.

A similar trend to the $N-O_W$ lifetimes is seen in the water-water lifetimes in figure 4.5. The lifetimes are all shorter than the pure water system lifetimes, and increase less rapidly with increasing t^*, suggesting that the presence of the glu molecules perturbates the system sufficiently to de-stabilise the water hydrogen bond. This may be due to the ability for water molecules to re-orientate faster than larger molecules [119]. Un-favourable energetic states, exerting forces on

all the molecules in the system are likely to re-orientate water molecules faster than glu molecules, resulting in water molecule bonds breaking quicker than in the pure water system.

Finally, we consider the bonds formed with the sodium ions in the glu simulations. Figure 4.6 shows the bond-lifetimes of the sodium ions in the CHARMM22/F3C system interacting with the two carboxyl oxygens of the glu-molecule and the oxygen of the water molecule. In each case there is again a steady increase in τ_{HB} with increasing t^*, although the increase for the $O_W - Na^+$ interaction is very small, indicating that water-sodium bonds do not generally break and re-form in a short time. The plots for the carboxyl-sodium ($O_C - Na^+$ and $O_{CE} - Na^+$) are similar, both increasing with t^*, indicating the presence of temporary breaks for both carboxyl groups. Similar trends were observed for the SPC/E-Glu and TIP3P-Glu simulations, but are not shown here.

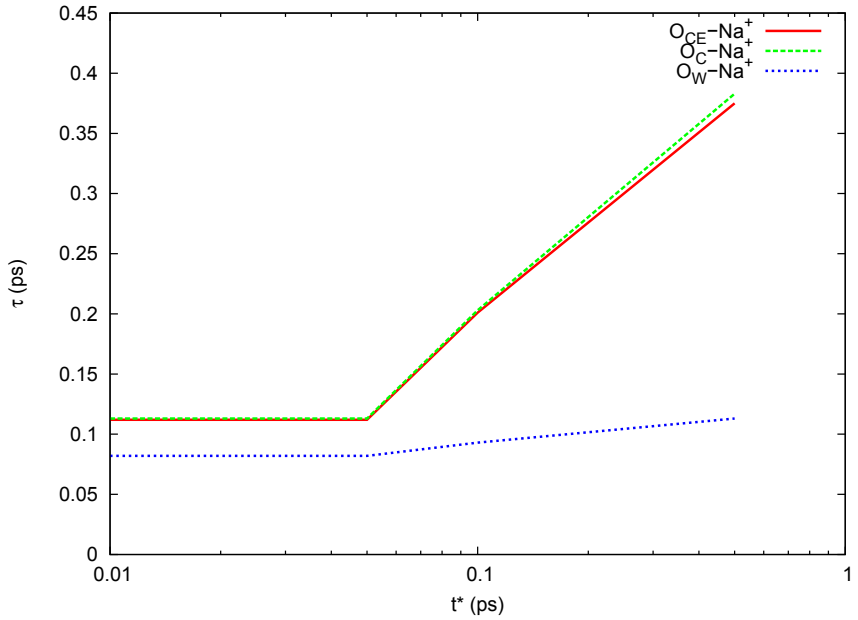


Figure 4.6: $\tau_{HB}(t^*)$ for the sodium-glu and sodium-water interactions in the F3C/CHARMM22 classical simulation of glu discused in chapter 3 for $t^* = 0.01$, 0.05, 0.1 and 0.5 ps.

The sodium-glu lifetimes are similar to the glu-glu lifetimes for $t^* = 0$, however, the glu-glu increase faster with t^* than the sodium-glu lifetimes suggesting that the sodium-glu lifetimes while susceptible to temporary breaks, are not as dependent on the value of t^* as the glu-glu lifetimes.

Figure 4.7 shows the relative bond lifetime of the sodium-glu and sodium-water interactions relative to the pure-water system lifetimes. These show that the $O_{CE} - Na^+$ and $O_C - Na^+$ bonds are longer lived than the pure-water bonds, and they increase faster with t^* than the pure-water bonds. This is, to some extent, expected as the bonds involving sodium are subject to a single radial cut-off to define the bond as only two atoms are involved. However, the sodium-water ionic-bond lifetimes are shorter than the pure-water hydrogen bond lifetimes, indicating that the sodium ions can move more freely than water in both the pure and aqueous solution.

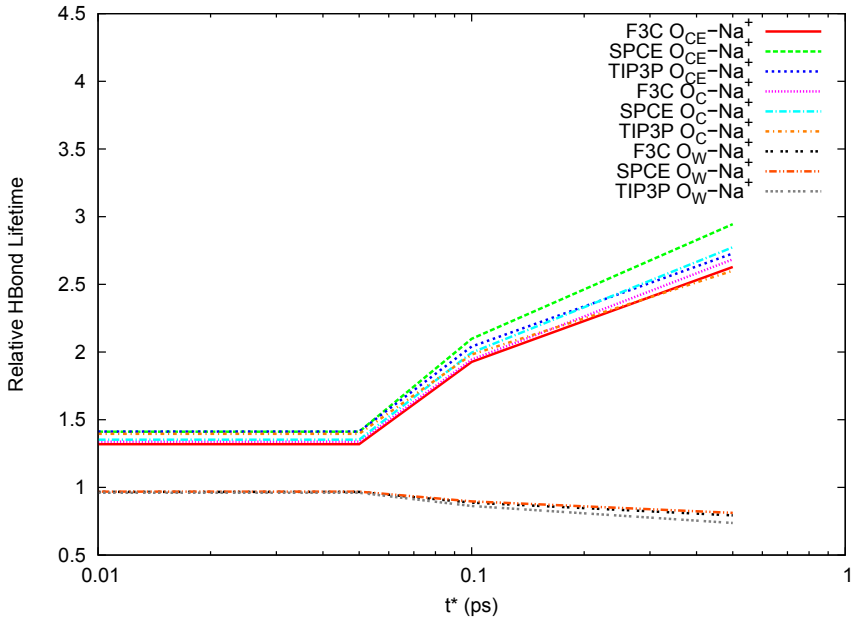
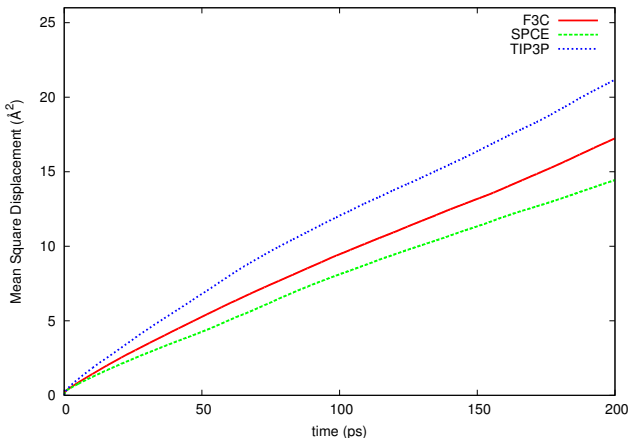


Figure 4.7: The relative bond lifetime, $\frac{\tau_{HB}}{\tau_{pure}}$ for the sodium-glu and sodium-water bonds, relative to the pure water system $O_W - H_W \cdots O_W$ hydrogen bond lifetime, for $t^* = 0.01$, 0.05, 0.1 and 0.5 ps.

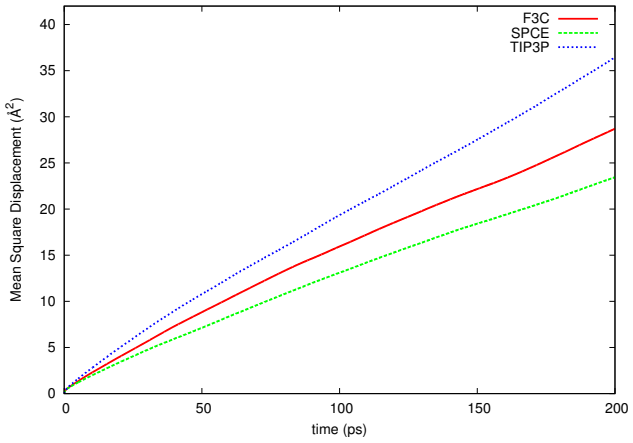
4.1.2 Mean squared displacement and diffusion

Figure 4.8 shows the mean squared displacement, $\langle |r(t) - r(0)|^2 \rangle$, of the glu molecules, sodium ion and water molecules in each of the three glu-solutions. The corresponding diffusion coefficients are shown in table 4.1.

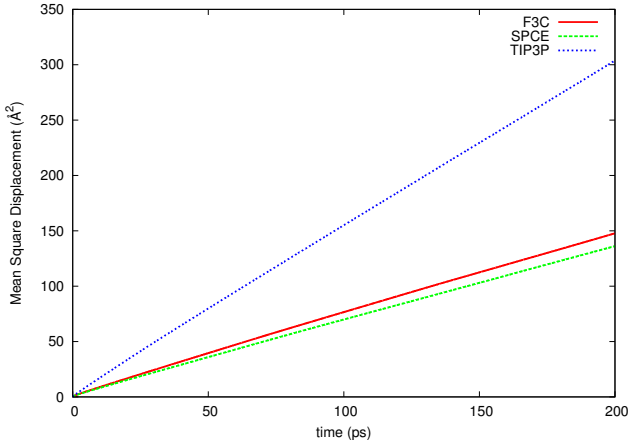
The mean squared displacements all increase fastest in the TIP3P system, indicating that the TIP3P water potential enables greatest freedom of movement of the molecules within the solution. The different axis limits of the $\langle |r(t) - r(0)|^2 \rangle$



(a) glu



(b) Sodium



(c) Water

Figure 4.8: Mean squared displacement of the glu molecules (4.8(a)), sodium ions (4.8(b)) and water molecules (4.8(c)) for each of the three glu simulations, Glu-F3C (F3C), Glu-SPC/E (SPCE) and Glu-TIP3P (TIP3P).

plots in figure 4.8 should be noted, as the glu and sodium ion plots (figures 4.8(a) and 4.8(b) are an order of magnitude smaller than the y-axis of the water plot (figure 4.8(c)). Water molecules are smaller than the glu-molecules and therefore it is not surprising that they diffuse faster. Despite the sodium ions being smaller than the glutamate and water molecules, the sodium ions diffuse only marginally faster than glu-molecules, and far slower than the water molecules, as shown by the diffusion coefficients in table 4.1. This indicates that the sodium ions are tightly bound to the surrounding molecules. The smaller self-diffusion of the sodium ions compared to water is also consistent with the longer bond-lifetimes compared to water shown in figure 4.7.

System	Molecule/Ion	Self diffusion (Å ps^{-1})
Glu F3C	glu	0.0133
Glu SPC/E	glu	0.0087
Glu TIP3P	glu	0.0149
Glu F3C	Na^+	0.0204
Glu SPC/E	Na^+	0.0137
Glu TIP3P	Na^+	0.0300
Glu F3C	Water	0.1222
Glu SPC/E	Water	0.1106
Glu TIP3P	Water	0.2478
Pure F3C	Water	0.0292
Pure SPC/E	Water	0.0254
Pure TIP3P	Water	0.0542

Table 4.1: Diffusion coefficients obtained from mean squared displacements of the glu, sodium ions and water in the three glu systems, and water in the three pure water systems.

The water molecules also diffuse more slowly in the pure-water system than in the glu solution. This is again consistent with the longer water-water hydrogen bond lifetimes in the pure-water system. This is somewhat surprising as Nieto-Draghi *et al.* found that the addition of methanol to water reduced the water-water hydrogen-bond lifetime [128]. In addition, Allison found that the addition of *N*-methlyacetamide (NMA) to water also decreased the water self-diffusion, although a slight increase in the self-diffusion was observed when increasing the NMA mole-fraction in water from 0.7 to 0.9 [27]. As this trend is observed for all three water potentials, this effect is not due to the parameteristion of

the water molecules and thus we conclude that it is due to the presence of the three charged groups of glu-molecules. The charged groups allow for strong glu-glu bonds to form. As noticed in the structural study in Chapter 3, glu-glu bonds are often stabilised by the presence of bifurcated bonds, suggesting that glu-glu bonds are not strong enough to overcome the forces producing the re-orientation. This is evidenced by the shorter glu-glu bond lifetimes than for water-water molecules in the pure water systems. The re-orientation of the glu molecules will inevitably result in perturbation of the water structure, as suggested by the lack of tetrahedral structure and poorly defined outer co-ordination shells in the study of water in the glu solution (see section 3.3.3) and suggested by Padró *et al.* [119] in their study of alcohol molecules in solution with water.

The forced re-orientation of glu-molecules and the mobility of the relatively small water molecules results in relatively small self-diffusion coefficients for all species within the glutamate solutions. We conclude that one of the factors that may increase the speed orientation is the number of charged sites capable of forming hydrogen-bonds on a glu-molecule (3 sites, involving a total of 7 atoms) compared to methanol as studied by Nieto-Draghi (1 site with an O-H^-) or NMA (two charged sites with just one positive and one negatively charged atom) [128, 27]. The multiple hydrogen-bond sites, coupled with the presence of sodium ions produces a system where molecules are constantly being attracted to form bonds in multiple directions, resulting in a system of relatively short-lived bonds, and relatively fast self-diffusion coefficients.

4.1.3 Conclusions on dynamics of aqueous glu solutions

The glu-glu amine-carboxyl bond was found to be the longest lived hydrogen bond within the glu-systems. The side-chain (O_{CE}) and the C_α (O_C) carboxyl bond's with the amine are of similar longevity. The shortest lived bonds are in the water-water interactions within the glu-systems. These, along with the sodium-water and amine-water bonds are all shorter-lived than the pure-water bonds, suggesting that the long-lived glu-bonds result in the rest of the system being de-stabilised. We also find that the type of water-potential has a greater impact on the glu-glu interaction than the type of carboxyl involved. The longevity of the glu-glu hydrogen bonds relative to a pure-water simulation and the small self-diffusion indicates slow dynamics of the glu-molecule, and is consistent with the

structural dependence on starting configuration of the molecules in the system.

The dependence of the lifetimes on t^* also indicates that these systems are particularly susceptible to temporary bond-breakages. Whether these temporary bond-breakages actually break the 'real' hydrogen-bond is, perhaps, a subject that should be considered further. The difference between the bond lifetimes involving glu-molecules across the range of t^* is dramatic. The temporary breaking of bonds and subsequent re-formation needs to be considered in dynamical studies as the re-formation reduces the overall mobility of the molecules. In the cases where there is little dependence on t^*, the mobility of the molecules is unaffected. If the re-formation cases were not considered the lifetimes would not take into account the reduced mobility of the molecules and it is not sufficient to only consider the case where $t^* = 0$. This data does not provide an obvious candidate for a value of t^* that would be optimal in all cases, where the chosen value ensures bonds that are temporarily broken are included as one long-lived bond, and it does not include bonds that are truly two separate motifs. Far more systems would need to be considered, with a broader spread of t^* values before any conclusions could be drawn.

The relatively long self-diffusion times for the sodium ions compared to both glutamate and water molecules may be relevant in the context of the glutamate transport by sodium ions in a biological environment. If the sodium ions are slow moving in this system, then their ability to move freely through a high concentration of aqueous glutamate to attach to glutamate that currently forms no sodium bonds would be slower than the speed of the glutamate moving. This may inhibit the use of sodium ions to transport glutamate, particularly when their is a extracellular glutamate build up such as in brain hypoxia [133]. Further study should be considered to find out if slow self-diffusion of the sodium ions does inhibit glutamate transfer, particularly with respect to the speed of glutamate release by the neurones. Other ions may behave differently in this environment, providing the long ion-glutamate lifetimes that we have identified here, but otherwise propagating fast through the environment, although as the propagation may be a result of the long-lived ion-glu bond this may not be possible.

These results emphasise the dependence of the structural information obtained in short-time Molecular Dynamics simulations on the initial configuration, as sug-

gested in chapter 3. Investigation of the structure of the glu solution *via* enhanced sampling techniques is essential to further our understanding and so in the following section we consider parallel tempering simulations of the glu solution.

4.2 Parallel Tempering Simulations of Glutamic Acid Solutions

Analysis of the structure of complex molecular systems is limited by the ability to efficiently sample the conformational space of the system under consideration. Complex systems such as the glu solution often have high-energy barriers between low-energy configurations. Due to the short-time scales and limited coverage of the available energy landscape in classical simulations, effective sampling of the possible configurations of such systems is limited. As noted in the classical studies in chapter 3, and from the dynamical behaviour reported in section 4.1, glu molecules diffuse relatively slowly. Parallel tempering is used as an enhanced sampling technique to obtain a clearer picture of the system structure, by simulating the same concentration system of glu:sodium:water at 1:1:29 with the PINY_MD package [109].

4.2.1 Methods

A single glu:sodium:water box containing 20 glu molecules, 20 sodium ions and 580 water molecules (using the SPC/E water geometry) was constructed using the same methods as described in the set-up of the classical simulation in chapter 3. We used a conjugate-gradient energy-minimisation algorithm for 10,000 steps to further optimise the geometry of the box with respect to the CHARMM22 and SPC/E forcefields. The system was then subjected to a 20 ps NPT simulation, at atmospheric pressure (101,325 Pa), 300 K, and with a 1 fs time-step, to enable equilibration of the box size. Once the correct box-size, and therefore density of the system was obtained, this was followed by a further 1 ns NVT equilibration procedure at 300 K, with 1 fs time-steps and using the Nosé-Hoover thermostat, with a Nosé-Hoover chain of length 2 [79].

Once a fully equilibrated system was obtained it was replicated n-times for the n-temperers (or replicas) needed for the parallel tempering simulation. In

parallel tempering a temperer is one copy of the system, simulated at a particular temperature, but able to exchange configuration with other temperers subject to a given probability.

Choosing the number of replicas required to accurately sample the conformational space has been the subject of much debate in recent literature [105, 107, 108, 139, 140, 141]. There are two key issues, as discussed in the methodology in section 2.2.4; firstly ensuring that the highest temperature enables the system to 'traverse' the high-energy barriers, and secondly, obtaining a temperature set between the highest and lowest temperatures that enables efficient exhange between the different replicas. The following section discusses these issues for the application of parallel tempering to the aqueous glu systems.

Choice of temperatures for different temperers

In chapter 3 it was identified that the clustering of the glu-molecule was dependent on the initial configuration of the system, and hence the distribution of cluster sizes was chosen as a 'measure' of the sampling of the conformational landscape of the system. The upper temperature for the temperature set needed to be chosen such that it enabled the transfer between small and large clusters of glu-molecules.

Initial simulations of the 1:1:29 glu system at 300 K, 400 K, 450 K and 500 K were performed. In each instance we used the PINY_MD package, with the system configuration that we had obtained for starting the parallel-tempering simulations, applying the Nosé-Hoover NVT thermostat, with a chain length 2, a 1 fs time-step and a total simulation time of 2 ns. From these simulations we calculated the distribution of cluster sizes (including 'clusters' containing a single molecule), as done in chapter 3. Figure 4.9 shows the distribution of cluster sizes at each temperature, averaged across the entire simulation.

The distribution of cluster sizes shows that there is a clear shift towards larger cluster sizes at 400 K compared to 300 K, although only the 450 K and 500 K simulations obtain cluster distributions as high as 17 members. As the difference between the 450 K and 500 K simulations is minimal, higher temperatures are not considered, particularly as we found in the artifically aggregated simulation considered in chapter 3 resulted in clusters of 17 or 18 members with single glu-molecules surrounding the aggregate. None of the simulations considered here obtain a 20-member cluster (the maximum cluster size as there are 20 glu

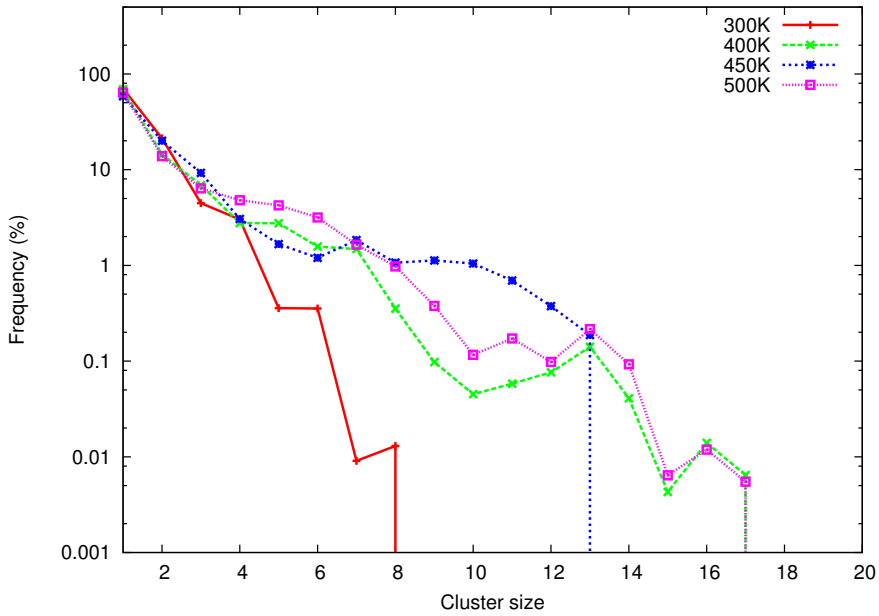


Figure 4.9: Plot showing the distribution of cluster sizes of an aqueous glu solution at different temperatures.

molecules in the system). However, the 500 K simulation has a larger proportion of clusters of 8 to 13 members than the 450 K simulation. As this study is concerned with not just obtaining larger clusters, but also characterising different types of small clusters, a top temperature of 500 K was used in the parallel tempering simulations. A top temperature of 450 K may have been adequate, but, to ensure that the sampling of the energy landscape in the short simulation time was effective as possible it was felt that the 500 K top temperature should be used.

The temperature distribution

The choice of temperature set should take into consideration the sets ability to efficiently exhange temperers during the parallel tempering simulation. The exchange of temperer configurations facilitates the enhanced sampling of this technique, as configurations that have been exposed to high-temperatures, and therefore able to traverse the energy landscape quickest, are then passed down to the lower temperatures.

Two temperers, i and j, exchange configurations subject to the following

probability;

$$P_{ij} = min\left\{1, \frac{exp(\Delta\beta_{ij}\Delta E_{ij})}{1 - exp(\Delta\beta_{ij}\Delta E_{ij})}\right\}, \tag{4.2}$$

where $\Delta\beta_{ij}$ is the difference in $\frac{1}{kT_i}$ and $\frac{1}{kT_j}$, and ΔE_{ij} is the difference in the total energy between the two replicas, E_i and E_j. If the replicas are too far apart then the configurations are less likely to be exchanged.

The subject of this work is not to study the optimal spacing of the temperature set. A geometric progression was chosen to span the temperature space, as Predescu *et al.* argue that this is the best initial distribution to use for parallel tempering simulations [142]. Starting with 25 temperers an initial parallel tempering simulation was performed, using 1 fs time-steps and attempting swaps every 50 time-steps (50 fs). After the first 44 swap-steps (2.21 ps) the temperatures of the set were adjusted to increase the probability of temperer-swap. From the initial 44 swaps, we calculated the acceptance ratio of the attempted swaps. If two temperers had a low probability of exchange we decreased the temperature difference between the two replicas, and in some instances this was achieved by adding a temperer to the temperature set. This was repeated a further 5 times, in each instance after 44 swap steps between each modification of the temperature set, and resulted in a total of 32 temperers each with an acceptance ratio between 15 and 25%. This was done because previous work has suggested an optimal distribution is achieved when the acceptance of swaps is 20% [103].

This 'optimised' set was then used for a 1.54 ns (30,800 swap steps) simulation. The temperature distribution used is shown in appendix B, table B.3.

4.2.2 Identification of conformations for further analysis

Parallel tempering implies that temperers traverse the temperature set*via*a random walk. For effective parallel tempering, at the end of the simulation several 'coastals', that is temperers that have traversed the entire temperature space from top to bottom or vice-versa, need to exist.

Analysis of the trajectories passing through the temperature space during the parallel tempering simulations identified 179 instances where the replica at the bottom temperature had originated from the very highest temperature, and 105 instances where the temperer at the top had originated from the very lowest

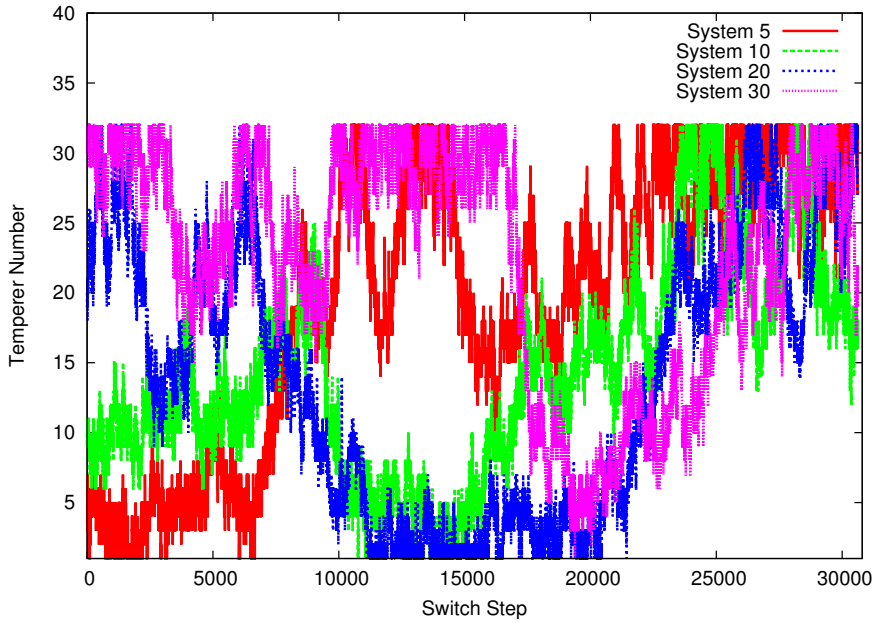


Figure 4.10: Plot of the movements of replicas 5 (red), 10 (green), 20 (blue) and 30 (pink) through the 32 temperers of the parallel tempering temperature set (from 300K to 500K), over the 30800 switch steps completed.

temperature. In comparison, Tulip *et al.* obtained just 6 coastals, 2 'down' that reached the bottom temperature from the highest, and 4 'up' [16], although this is for a much larger system, and covering a greater temperature range. The comparitively high number of coastals indicates that this parallel tempering study allows for sufficient mobility through the temperature set, thereby increasing the sampling of the energetic landscape. Figure 4.10 illustrates how a small selection of the replicas traversed the temperature space, illustrating the random walk.

Once the efficacy of the parallel tempering simulations, in terms of traversal of the temperature set, had been assessed, we considered the conformations of the system at the end of the full 30,800 attempted swap-steps. As one of the aims was to change the clustering behaviour, this was used as a tool for identifying new motifs that would benefit from further analysis. We calculated the distribution of cluster sizes for each of the 32 temperers for the last 2 ps of the parallel tempering simulation; these are shown in figure 4.11. This enabled the identification of conformations that were representative of all 32 conformations obtained from the parallel tempering. From this we were able to group the replicas as shown in table 4.2. The final configurations of replicas 12, 21, 22 and 32 provide distinctly

different clustering motifs and therefore these are considered in further study. These are used as the initial configurations in the production runs discussed in the rest of this chapter.

Replicas	Description
11-17	Similar clustering between 6 and 14 member cluster sizes. Replica 12 has clustering throughout this region.
21, 23, 25	Clusters ranging from 6-17 members with similar propensity. Replica 21 has largest cluster sizes at 17 members, as well as greatest intensity at size 6.
22, 24	Similar distribution to that of replicas 21 and 23. However, these replicas have 14, and 16 member-clusters, which replicas 21 and 23 do not have .
1-10,19-20 26-32	Clustering distirbution that is between the initial-state and the clustering observed for replicas 11-17. Replica 32 is representative with non-zero freqency for clusters of 1 and 6 members, with peaks at 8 and 10.

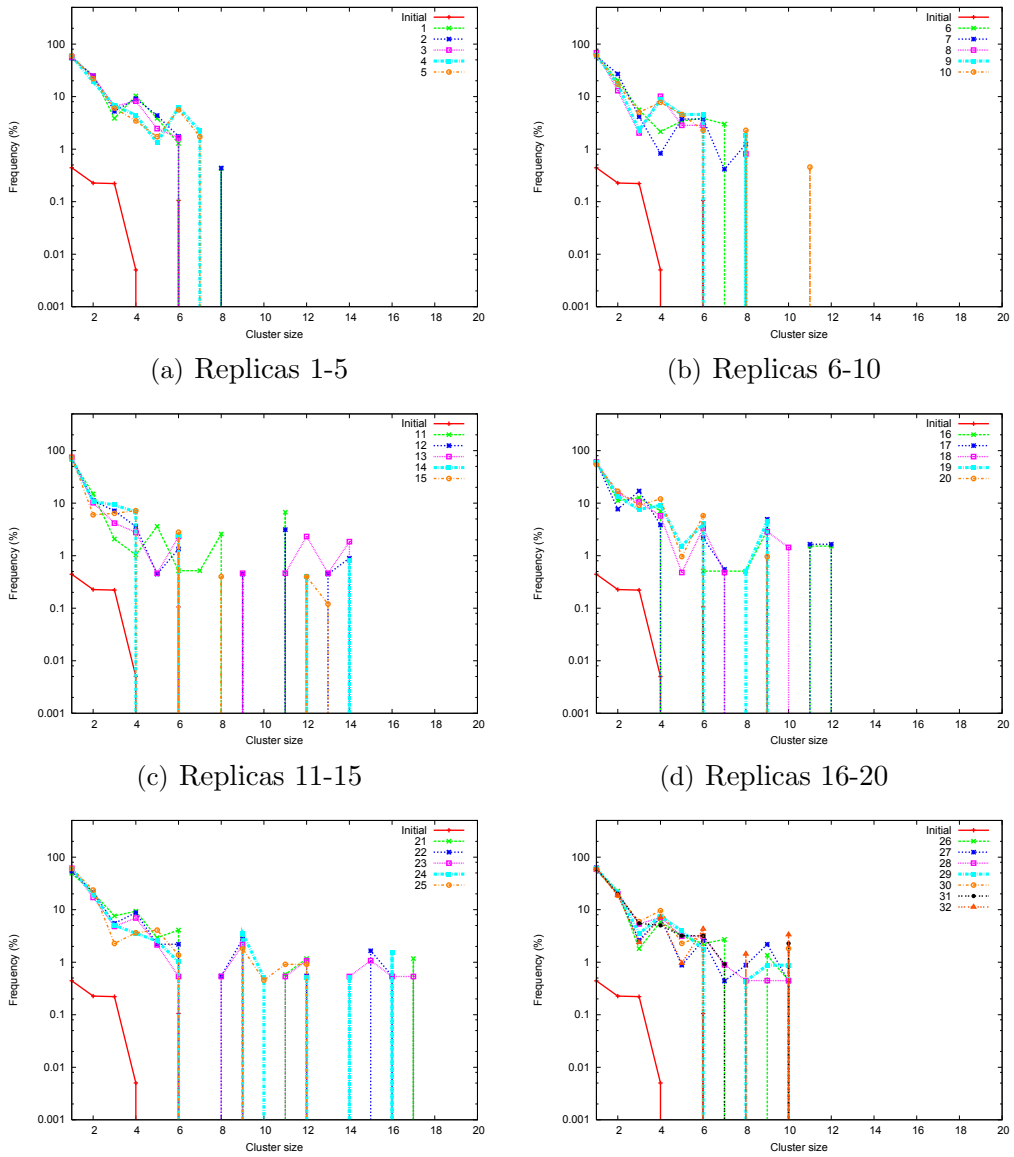
Table 4.2: Summary of the key features of the cluster distribution graphs of figure 4.11. We have grouped similar distributions together to enable choosing representative systems to cover all motifs achieved *via* the parallel tempering work.

4.2.3 Summary of simulations

To obtain the production simulations that are discussed in the following section, we carried out several steps to obtain the four unique initial configurations, as discussed above. This simulation scheme is summarised in the flowchart in figure 4.12.

For all the NVT simulations, including the parallel tempering, the Nosé-Hoover NVT thermostat, with a 1 fs time-step, and a Nosé-Hoover chain of length 2 were used. The parallel tempering involved 32 replicas, attempting swaps every 50 fs (50 time-steps) for a total total 1.54 ns simulation. The simulations were carried out in parallel using 32 processors (1 temperer per process) on the UK National Computing Facility, HPCx [130].

After the parallel tempering, five configurations were subjected to further simulation; the four unique configurations identified from the end of the parallel tempering simulation and the initial configuration that was used to start the



(a) Replicas 1-5

(b) Replicas 6-10

(c) Replicas 11-15

(d) Replicas 16-20

(e) Replicas 21-25

(f) Replicas 26-32

Figure 4.11: Cluster distribution for the last 440 sampled time-steps for each of the 32 replicas. These are also plotted against the distribution of glu cluster sizes for the initial structure used to start the parallel tempering simulation.

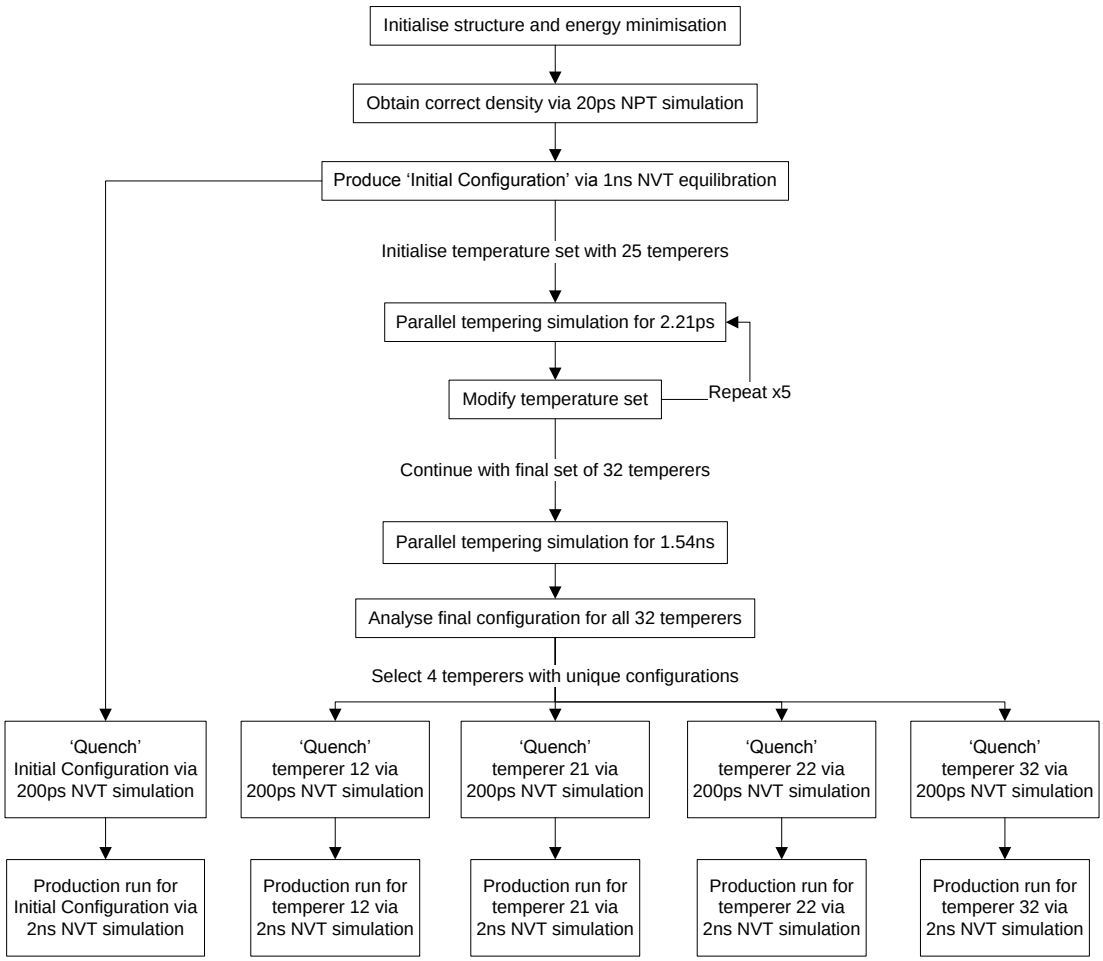


Figure 4.12: Schematic of the simulations carried out for the parallel tempering analysis.

parallel tempering. These were all subjected to a 200 ps NVT 'quench' simulation at 300 K, to enable the configurations from higher temperatures to obtain configurations more consistent with 300 K. This quench is needed to make the micro-structures more representative of those obtained at 300 K, however, it is short enough that it does not 'anneal' out the macrostructure obtained during the parallel tempering (this is evidence by the range of cluster sizes still present in the simulations considered in the results discussed in section 4.2.4). This was followed by a 2 ns production run, again using the Nosé-Hoover NVT thermostat at 300 K, with a 1 fs time-step, and a Nosé-Hoover chain of length 2. The trajectory was sampled every 100 fs, resulting in 20,000 configurations for analysis.

In the following discussion we refer to six systems:

- Original - the original simulation carried out in DLPOLY_2 and discussed in chapter 3. This is a 1 ns NVT simulation at 300 K, with a time-step of 0.5 fs.
- Initial - the initial configuration used to start the parallel tempering after a 1 ns NVT equilibration. This was then subjected to a 200 ps quench NVT simulation, and a 2 ns production run, at 300 K, with a 1 fs time-step.
- Temperer 12 - the final configuration of replica 12 at the end of the parallel tempering simulation. This was then subjected to a 200 ps quench NVT simulation, and a 2 ns production run, at 300 K, with a 1 fs time-step.
- Temperer 21 - the final configuration of replica 21 at the end of the parallel tempering simulation. This was then subjected to a 200 ps quench NVT simulation, and a 2 ns production run, at 300 K, with a 1 fs time-step.
- Temperer 22 - the final configuration of replica 22 at the end of the parallel tempering simulation. This was then subjected to a 200 ps quench NVT simulation, and a 2 ns production run, at 300 K, with a 1 fs time-step.
- Temperer 32 - the final configuration of replica 32 at the end of the parallel tempering simulation. This was then subjected to a 200 ps quench NVT simulation, and a 2 ns production run, at 300 K, with a 1 fs time-step.

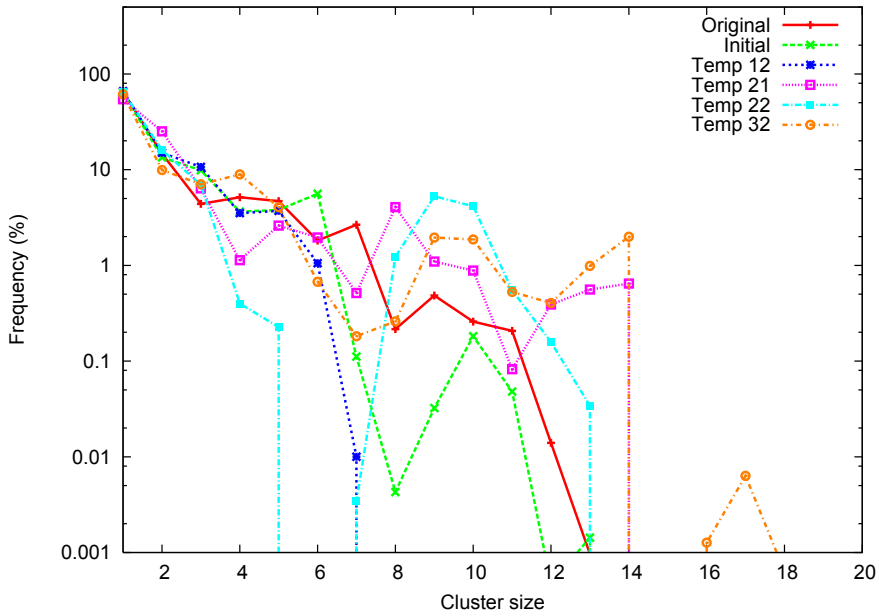


Figure 4.13: Distribution of cluster sizes averaged over entire 2 ns simulation post-parallel-tempering for each of the systems simulated. We also show the original cluster distribution for the CHARMM22-SPC/E simulations carried out using DLPOLY_2 for the classical results in chapter 3.

4.2.4 Glutamate structure after parallel tempering

Clustering of glutamate molecules

As one of the key issues with the initial classical simulations of glu considered in chapter 3 was the lack of mobility between small and large clusters of glu molecules, we first consider the difference in cluster sizes in the new simulations.

Figure 4.13 shows the distribution of cluster sizes for each of the systems considered post-parallel-tempering. The plot shows the percentage of clusters of a particular size, as a percentage of the total number of clusters, including cluster sizes of 1 (a single glu molecule) and up to 20 (all the glu molecules in the system). The results for each system are the average from the entire 2 ns trajectory. For comparison purposes we also show the distribution of clusters obtained in the 'original' classical CHARMM22-SPC/E simulations carried out using DLPOLY_2 from chapter 3.

The plot shows that each system has maintained a similar cluster distribution to those obtained at the end of the parallel tempering simulations, resulting in the broad range of cluster sizes produced by the six systems. This is to be

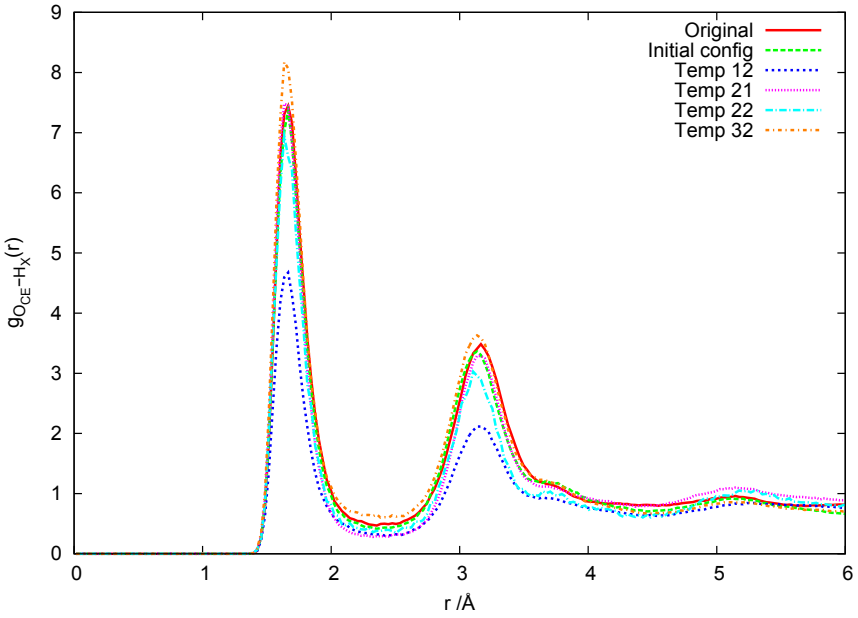
expected due to the slow changes observed in the initial simulations in chapter 3. In particular the temperer 32 system maintains clusters (albeit small numbers) of 16 and 17 glu molecules. We also see that the temperer 12 system maintains cluster sizes only up to a maximum of 7. The temperer 21 system maintains a constant range of cluster sizes between 1 and 14. The temperer 22 system has a spike in intensity of cluster sizes between 8 and 11 and the 'initial' system, which was not subjected to parallel tempering, gives a similar distribution to that of the original DLPOLY_2 simulation.

These results show that the five systems have maintained their different characteristics that were obtained *via* parallel tempering even after the 2 ns standard NVT simulations, supporting the evidence of a slow-moving system, with high energy barriers, and the need for better sampling techniques as provided by parallel tempering. This is only one structural property of the systems, but as this difference has been maintained in the production run, any other features that the enhanced sampling has been able to obtain may also still be present. Next we consider the structural properties of the intermolecular interactions of the system.

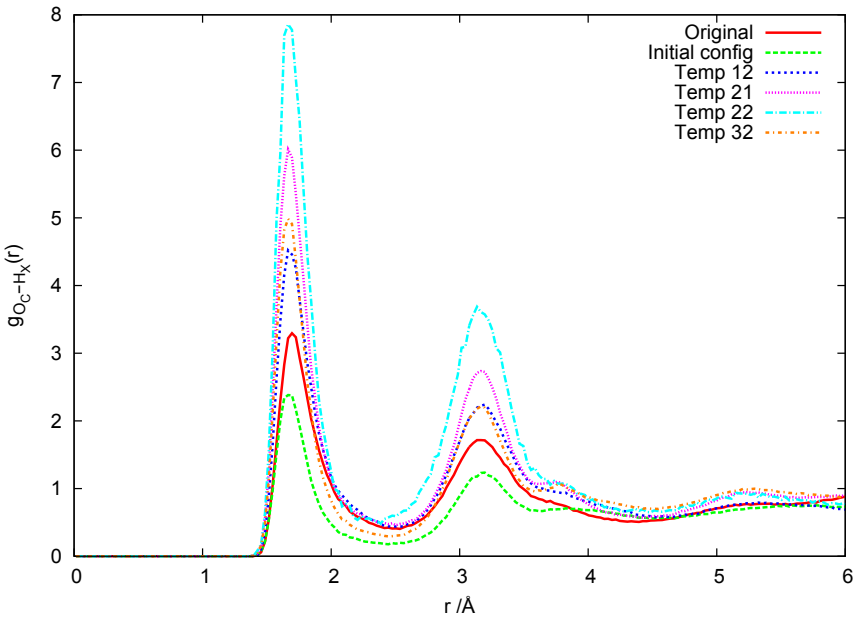
Glutamate-glutamate interactions

Figure 4.14 shows the radial distribution functions (RDFs) for the interactions between glu molecules *via* the amine H_X atoms and the two species of carboxyl oxygen, O_C and O_{CE} (the C_α and side-chain carboxyl oxygens respectively) for the four post-parallel-tempering simulations, the initial configuration simulation and the original results obtained using DLPOLY_2 in chapter 3.

The RDFs for the $O_C - H_X$ interaction, figure 4.14(b), show a large variation in the intensities of the first and second peaks across the different systems. There is less variation than this in the $O_{CE} - H_X$ plot, figure 4.14(a), particularly in the second peak. The greatest intensity in the $O_C - H_X$ plot is in the temperer 22 system, which has a first-peak intensity nearly four times that of the smallest first peak in the same plot. The second peak is similarly more intense, which is to be expected as the first peak is due to the hydrogen bond contact $H_X \cdots O_C$, and the second peak is due to the second oxygen on the same carboxyl group co-ordinating with the H_X atom. The greater intensity of these peaks suggests that the temperer 22 system has a greater proportion of $O_C - H_X$ interactions than the



(a) $O_{CE} - H_X$



(b) $O_C - H_X$

Figure 4.14: Radial distribution functions of the glu-glu interactions from the initial configuration system, the four post-parallel-tempering simulations and the original simulation.

other systems, and may also indicate larger glu-clusters as large clusters would require a greater number the greatest number of glu-glu bonds to be formed. The clustering distribution (figure 4.13) shows that the temperer 22 system has clusters of up to 13 members, but no larger.

The second highest first-peak in the $O_C - H_X$ RDF is from the temperer 21 system, which had clusters with up to 14 members. The smallest peak is from the initial-configuration system. This is consistent with the clustering distribution as the original and the initial configuration system exhibited the smallest proportion of large clusters, with no clusters larger than 12, and only a very small proportion above 8 members.

There is also a difference in the intensities of the $O_{CE} - H_X$ RDFs in figure 4.14(a). This is reflected in the wide range of $H_X - O_{CE}$ co-ordination numbers shown in table 4.3, ranging from 0.4 to 0.9 (the original system had a co-ordination number 0f 3.2). The difference in the $H_X - O_C$ co-oridination numbers is even more marked, ranging from 0.3 to 1.3 (the original simulation had a co-ordination number of 0.37). This indicates that there is a great variation in glu-glu bond distributions depending on the current configuration, and thereby highlights the importance of the parallel tempering simulations. The initial configuration system and temperer 12 both form more glu-glu bonds with the side chain, whereas the other systems form the majority of the glu-glu bonds *via* the C_α-caboxyl, indicating that this again depends on the system configuration currently available to the simulation.

System	$n_{O_{CE}}^{H_X}$	$n_{O_C}^{H_X}$
Original F3C	0.32	0.37
Initial	0.8	0.3
Temp 12	0.9	0.5
Temp 21	0.7	0.7
Temp 22	0.8	1.3
Temp 32	0.4	0.5

Table 4.3: Co-ordination numbers n_α^β at a radial distance $r_{min} = 2.4$ for the two species of glu-glu bonds for each post-parallel tempering simulation and the original system discussed in chapter 3.

Of further interest is the presence of a shoulder in the trailing edge of the second-peak in both plots (between 3 and 4 Å). This shoulder is present in the

original data (the yellow line in figures 4.14(a) and 4.14(b)), however, it was not as clear and therefore was not discussed as a significant motif in the original classical study in chapter 3. In the post-parallel-tempering RDFs it is more prevalent, particularly in the $O_C - H_X$ RDF. One possible explanation for this feature is the presence of the second and third hydrogen atoms in the amine group (H_X), that are not directly bonded with one of the carboxyl oxygens, co-ordinating with the un-bonded carboxyl oxygen. That is if O_{C1} and H_{X1} form a bond, then O_{C2} co-ordinates with H_{X2} and H_{X3} in a non-bonded fashion.

The shoulder in the second peak exists in the region 3.6 Å $\leq r_{O_{CE}/O_C-H_X} \leq$ 4.4 Å. From visual observation of the trajectory using the VMD molecular visualisation package [125], we were able to identify motifs between H_X and O_{CE}/O_C atoms that correspond to this shoulder. Figure 4.15 shows a diagram of the amine and C_α-carboxyl groups from two different glu molecules interacting. This shows that for a bonded $O_{C1}\cdot\cdot H_{X1}$ interaction, with the two species of atom approximately 1.6 Å apart (consistent with the first peak in the g(r_{O_{CE}/O_C-H_X}) RDFs), that there is an interaction between the unbonded O_C atom (O_{C2}), and the second, unbonded, H_X atom (H_{X2}) of 3.1 Å, and between the bound O_C atom (O_{C1}) and H_{X2} which is 3.2 Å. Both of these distances are consistent with the second peak in the g(r_{O_{CE}/O_C-H_X}) RDFs. The final interaction between the non-bonded O_{C2} atom, and the third non-bonded H_X atom (H_{X3}), is at a distance of 3.9 Å. This is consistent with the position of the shoulder in the trailing edge of the second peaks in the g(r_{O_{CE}/O_C-H_X}) RDFs.

A question remains as to why the shoulder motif discussed here is more common in some of the post-parallel-tempering systems than the original solution and the initial-system, which was not subjected to parallel tempering, but was subjected to the same post-parallel-tempering simulation protocol. It is also more prevelant in the temperer 21, 22 and 32 systems than in the temperer 12 system. There is no clear explanation for the difference, particularly as the shoulder appears in all six simulations, just with differing intensities.

Figure 4.16 shows the RDFs for the interaction between the amine hydrogen atom, N, and each of the two species of carboxyl oxygen, O_{CE} and O_C. These provide additional evidence that glu-glu bonds prefer to form *via* the C_α-carboxyl and the amine group for the temperer 22 system, and for side-chain-carboxyl and amine interactions for the temperer 21 and 32 systems and the initial

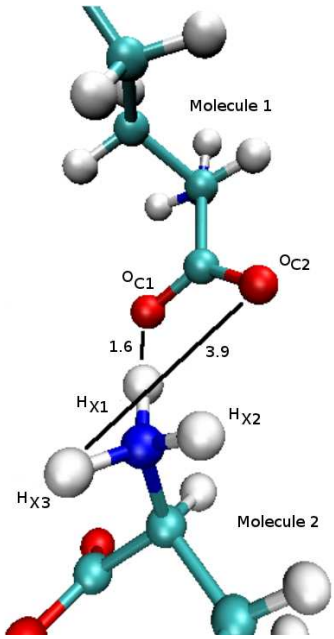


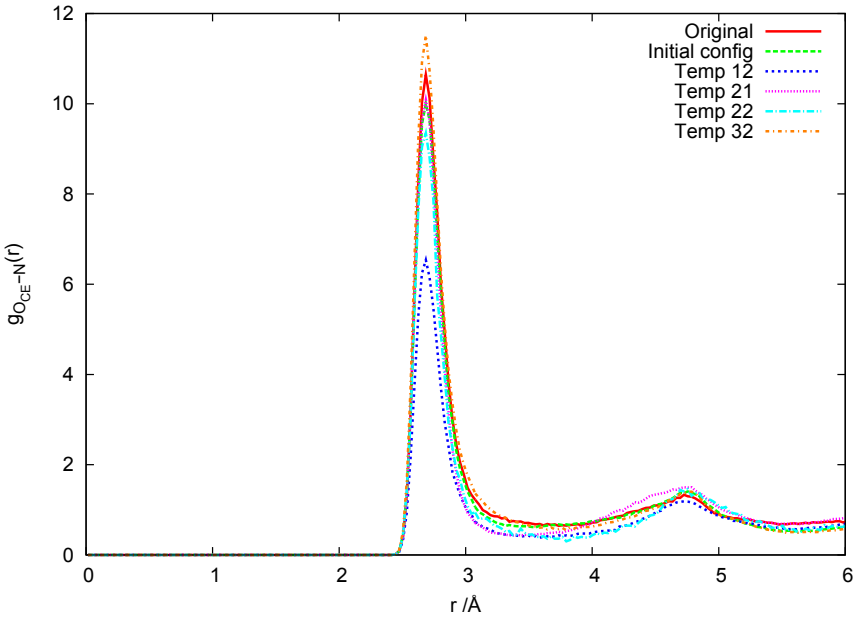
Figure 4.15: Schematic of glu-glu bonding motif corresponding to shoulder on the trailing edge of the second peak in the $O_{CE} - H_X$ and $O_C - H_X$ RDFs in figure 4.14 .

configuration system, with no clear preference in the temperer 12 system.

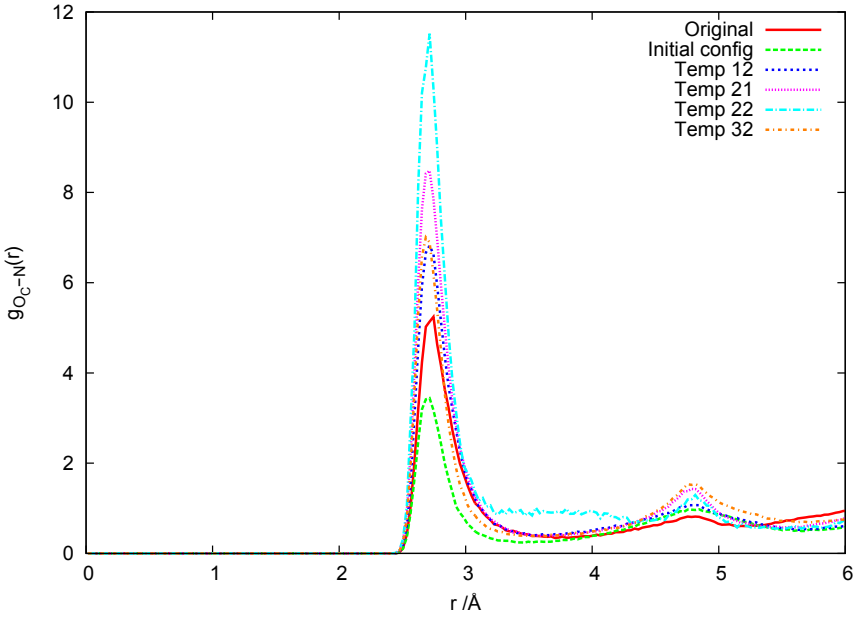
We next consider the 2D radial-radial distribution functions. Figure 4.17 plots the correlation between the $H_X - O_{CE1}$ and $H_X - O_{CE2}$ radial distances, where O_{CE1} and O_{CE2} are the two carboxyl-oxygen atoms from the side-chain group of the glu molecule. Figure 4.18 shows the similar 2D radial-radial distribution plots correlating the H_X and O_C atoms. For comparison we also plot the results from the original simulation as shown in chapter 3. The other four systems have similar plots, which are shown in Appendix B, figures B.1 and B.2.

We see similar results between the post-parallel-tempering simulations and the original data for the 2D radial-radial distribution plots, in both the $H_X - O_{CE}$ interaction (figure 4.17) and the $H_X - O_C$ interaction (figure 4.18). In each case there exist single linear bonds, indicated by the non-zero intensity in the regions $r_{H_X - O_{CE}/O_{C1}} \simeq 1.9$ Å, $2.2 \leq r_{H_X - O_{CE}/O_{C2}} \leq 3.9$ Å and $r_{H_X - O_{CE}/O_{C2}} \simeq 1.9$ Å, $2.2 \leq r_{H_X - O_{CE}/O_{C1}} \leq 3.9$ Å. There are also bifurcated bonds for all of the simulations, including the original system and the initial-configuration system, indicated by the non-zero intensity in the region $r_{H_X - O_{CE}/O_{C1}} \simeq r_{H_X - O_{CE}/O_{C2}} \simeq 2.0$ Å.

Despite the similarities in the plots, there is a difference in intensity between

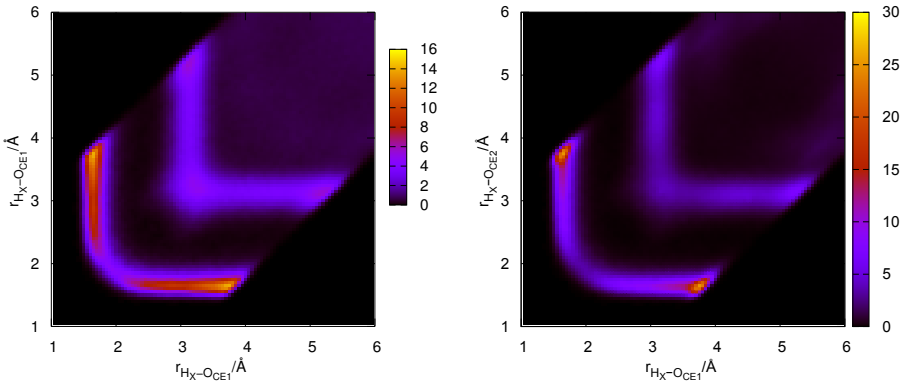


(a) $O_{CE} - N$



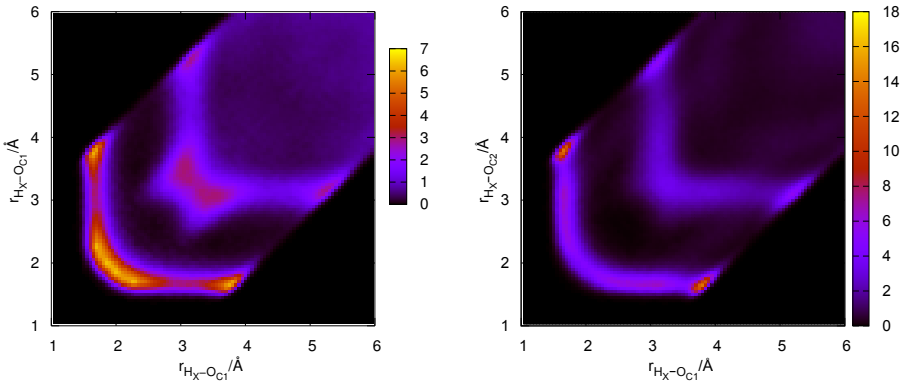
(b) $O_C - N$

Figure 4.16: Radial distribution functions of the glu-glu interactions from the original simulations of CHARMM22 glu and SPC/E water discussed in chapter 3, and the 5 post-parallel-tempering systems.



(a) Original system $H_X - O_{CE}$ (b) Temperer 22 $H_X - O_{CE}$

Figure 4.17: Two-dimensional radial-radial distribution functions of the glu-glu interactions between O_{CE} and H_X atoms.



(a) Original system $H_X - O_C$ (b) Temperer 22 $H_X - O_C$

Figure 4.18: Two-dimensional radial-radial distribution functions of the glu-glu interactions between O_C and H_X atoms.

the original and temperer 22 simulation plots. In the original classical simulations, discussed in the chapter 3, all three different systems simulated, using different water potentials, produced similar plots. The $H_X - O_{CE}$ 2D radial-radial distributions in figure 4.17(b) show a much greater preference to form bonds in the linear region ($r_{H_X-O_{CE1}} \simeq 1.9$ Å, $r_{H_X-O_{CE2}} \simeq 3.9$ Å and $r_{H_X-O_{CE2}} \simeq 1.9$ Å, $r_{H_X-O_{CE1}} \simeq 3.9$), than the bifurcated region. The original system (figure 4.17(a)) also shows a preference for linear bonds but it is far less extreme.

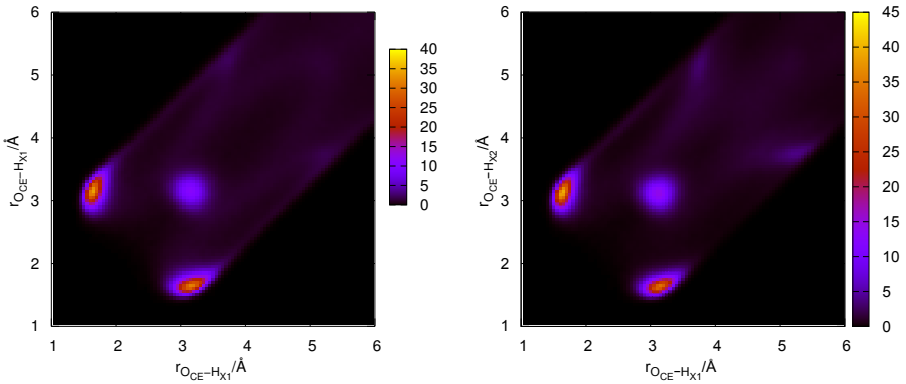
Similarly the H_X-O_C 2D radial-radial distribution for the temperer 22 system prefers to form linear rather than bifurcated motifs (figure 4.18(b)). In this instance the original system has a marked preference for bifurcated bonds over linear bonds (figure 4.18(a)). The temperer 21 results (figures B.1(d) and B.1(d) in appendix B) show similar distribution plots to that of temperer 22 for the $H_X - O_{CE}$ and $H_X - O_C$ interactions, favouring linear bonds over the bifurcated state. The temperer 32 plots suggest that glu-glu bonds favour single linear bonds on the side-chain (figure B.1(f)), but not on the C_α-carboxyl (figure B.2(f)). Whereas the initial configuration system (figures B.1(b) and B.2(b)) and the temperer 12 system (figures B.1(c) and B.1(c)) both show similar plots to the original system, although there is a less clear preference for $O_C - H_X$ bifurcated states.

These results indicate that despite evidence that the CHARMM22 empirical potential has a propensity to form more bifurcated states than other potentials [114], that there is clearly a difference in the favouring of bifurcated states depending on the exact conformation of the system. Bifurcated bonds still exist in the post-parallel tempering simulations, but as a smaller proportion of the total bonds, indicating that the reliance on potential can not be ruled out. The CPAIMD results also indicated the presence of a bifurcated state these results may call into question whether empirical force-fields do indeed always over-bifurcate and indicate that the problem is far more complex than this. In the future it may be prudent to examine enhanced-sampling of other apparently over-bifurcated simulations to assess whether similar results occur, as this may not be due to the CHARMM22 empirical potential.

Other than energetic impediments, the reason for the differing proportions of glu-glu bifurcated bonds between systems is unclear. If the glu-glu contact is unfavourable, then the bifurcated motif can provide a lower energy pathway

to a glu-glu contact. If single linear bonds are possible they are preferred as they are longer lived and stronger contacts. Presumably, the sampling technique provided by parallel tempering has allowed some of the systems to re-orientate their glu-glu interactions to favour linear bonds, without the need for bifurcated states as an intermediate step between linear bonds. If this is the case, these results further emphasise the slow dynamics of this glu system, and the re-orientational impediment this produces, and therefore the necessity to use an enhanced sampling technique such as parallel tempering.

As with the original system we also found that bifurcated states do not form where one O_{CE}/O_C atom is bound to two H_X atoms from the same amine. This is indicated by the zero intensity in the $r_{O_{CE}/O_C-H_{X1}} \simeq r_{O_{CE}/O_C-H_{X2}} \simeq 2.0$ Å in the 2D radial-radial distribution plots correlating the $r_{O_{CE}/O_C-H_{X1}}$ and $r_{O_{CE}/O_C-H_{X2}}$ distances, as shown in figure 4.19 and 4.20. Again we only show the original system and the temperer 22 system here. Similar results were obtained for the other four systems (figures B.3 and B.4).

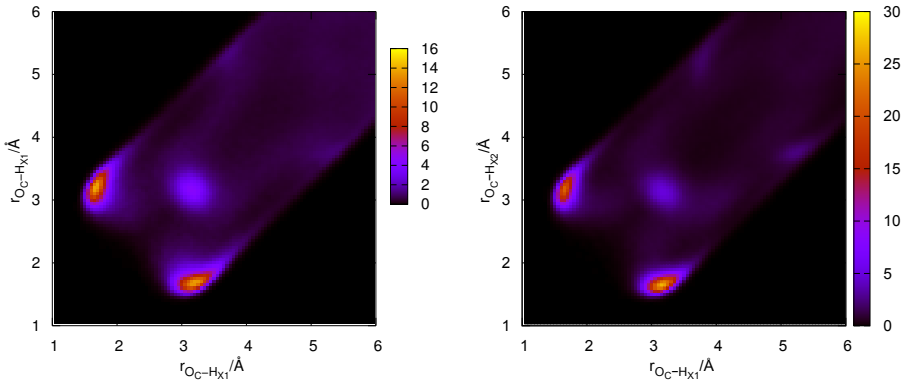


(a) Original system $O_{CE} - H_X$ (b) Temperer 22 $O_{CE} - H_X$

Figure 4.19: Two-dimensional radial-radial distribution functions of the glu-glu interactions between O_{CE} and H_X atoms.

Glutamate-water Interactions

The radial distribution functions between the carboxyl groups of glu molecules and the oxygen atoms of water molecules (O_W) from the post-parallel-tempering simulations and the original SPC/E simulations of chapter 3 are plotted in figure 4.21. Figure 4.22 shows the corresponding carboxyl-H_W (water-hydrogen) interaction. For both the carboxyl-O_W and the carboxyl-H_W interactions, the



(a) Original system $O_C - H_X$ (b) Temperer 22 $O_C - H_X$

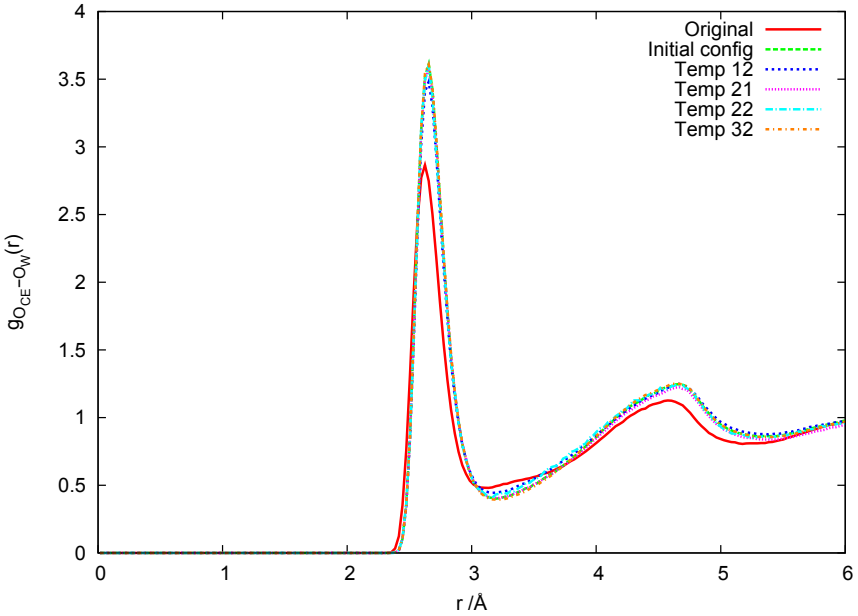
Figure 4.20: Two-dimensional radial-radial distribution functions of the glu-glu interactions between O_C and H_X atoms.

five post-parallel-tempering simulations closely follow the original system RDFs. The intensity of the first peak in all four plots is greater for all of the post-parallel-tempering systems, including the initial-configuration system which has not undergone the enhanced sampling of the parallel tempering simulation. This increase in intensity indicates more glu-water bonds forming than in the original simulation.

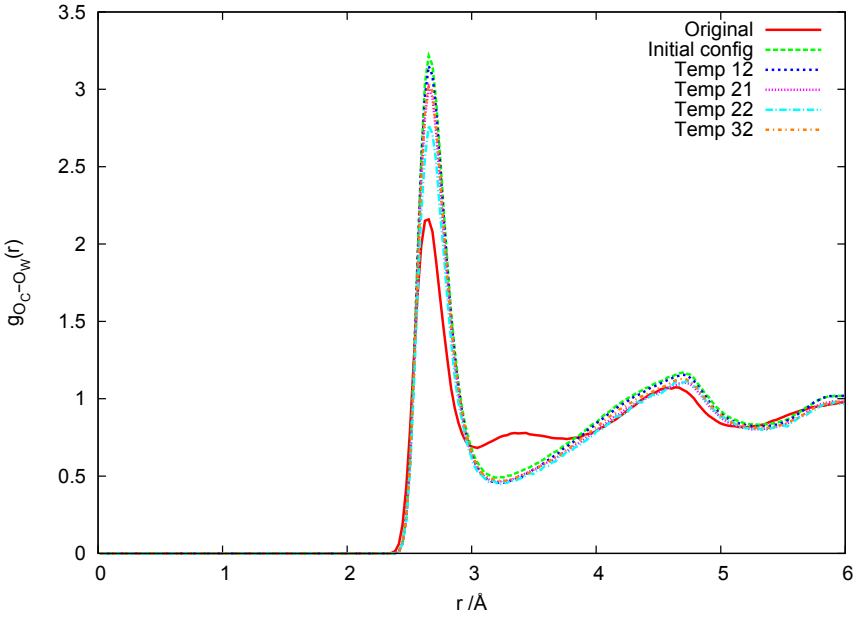
In the original simulations we found that all three water models produced similar intensity first-peaks in the $O_{CE}/O_C - O_W/H_W$ RDFs. Here we see a small difference that may be due to the length of the simulation, but the co-ordination numbers show very similar hydration of the two types of carboxyl oxygens by water molecules (table 4.4). The $O_{CE} - O_W$ occupation number is consistently greater than the $O_C - O_W$ occupation number for all five systems, but the difference is minimal as is the difference across the systems.

System	$n_{O_W}^{O_{CE}}$	$n_{O_W}^{O_C}$
Original	3.18	3.13
Initial	3.14	3.04
Temp 12	3.07	2.96
Temp 21	3.12	2.85
Temp 22	3.13	2.90
Temp 32	3.14	2.89

Table 4.4: Co-ordination numbers n_α^β at a radial distance $r_{min} = 3.2$ for the glu-water $O_{CE} - O_W$ and $O_C - O_W$ interactions .

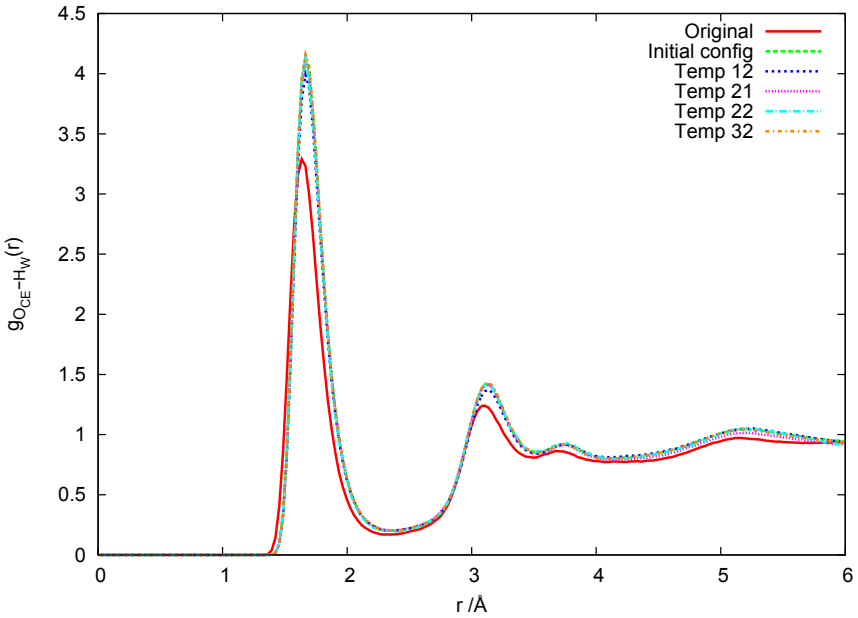


(a) $O_{CE} - O_W$

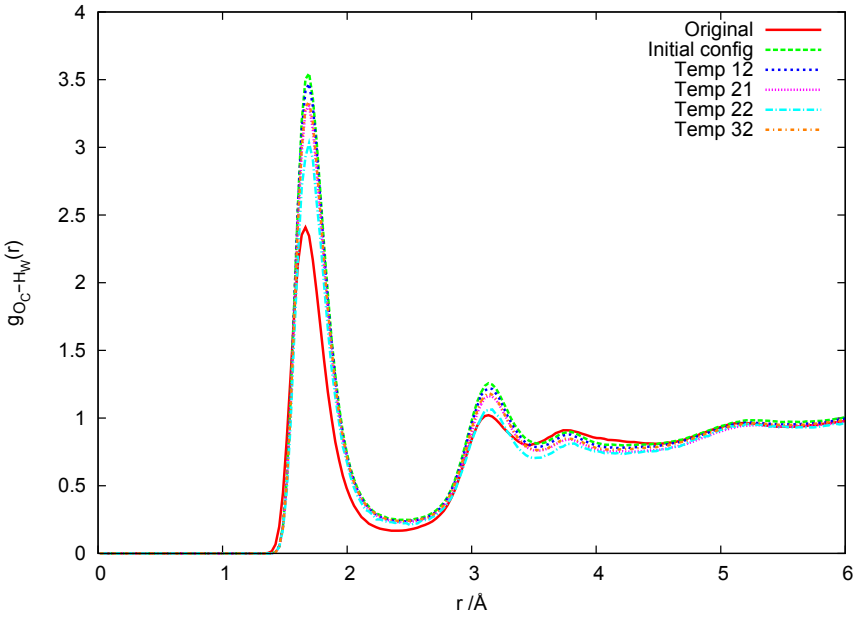


(b) $O_C - O_W$

Figure 4.21: Radial distribution functions of the glu-water carboxyl-O_W interactions from the original simulations of CHARMM22 glu and SPC/E water discussed in chapter 3, and the post-parallel-tempering simulations.



(a) $O_{CE} - H_W$



(b) $O_C - H_W$

Figure 4.22: Radial distribution functions of the glu-water carboxyl-H_W interactions from the original simulations of CHARMM22 glu and SPC/E water discussed in chapter 3, and the post-parallel-tempering simulations.

Of further interest is the lack of the intermediate peak between the main first and second peaks in the $O_C - O_W$ RDF (figure 4.21(b)). This is absent from all of the new simulations, including the initial-configuration system that has not been subjected to parallel tempering, but is present in the original simulation. In the original simulations this was attributed to the 'circle'-motif where a water molecule bonds to the C_α-carboxyl and the amine group of the same glu-molecule (this is discussed in more detail in chapter 3, section 3.3.3). Daub *et al.* also identified a similar peak in their glu-water RDFs from their classical study, although they attributed it to sodium ions mediating the water-carboxyl interaction [45].

As these simulations all use the same empirical potential for the glu and water molecules, and in addition Daub *et al.* had a similar peak in their RDFs, which uses the OPLS/AA force-field for glu molecules, this suggests that the absence of this peak in the new simulations is not an artefact of using a different simulation package (PINY_MD compared to DLPOLY_2 for the original simulations). We also found an intermediate peak, although it appeared more as an individual peak than a shoulder, but at the correct radial distance, in the Car-Parrinello simulation results (see chapter 3, figure 3.24(b)).

Analysis of the temperer 22 system for the 'circle'-motif, using the same criteria as described for the classical results in chapter 3, found a total of 33,112 circle motifs, accounting for just 0.008% of the $O_W - O_C$ bonds found. In contrast 2% of the $O_W - O_C$ bonds formed 'circle'-motifs in the original SPC/E simulation. The absence of the shoulder in the new simulation RDFs is most likely due to the differing proportions of circle motifs in the PINY_MD simulations. The absence of the shoulder on the first peak from *all* of the post-parallel-tempering systems is puzzling, particularly as these simulations use the SPC/E water potential and in the original simulations the motif was most prevalent in the SPC/E simulation (see chapter 3, figure 3.12(b)). This difference will be re-visited later in the discussion.

Figure 4.23 shows the RDFs for amine-water interactions. As with the carboxyl-water RDFs we see good agreement between all six systems in the position and intensity of the peaks, indicating that the original, initial-configuration and post-parallel-tempering systems all produce water-amine bonds of a similar nature. Again, the original simulation has a consistently smaller first-peak, and

a lower co-ordination number than the the post-parallel-tempering systems (see table 4.5), indicating more water-glu bonds forming on the amine site of the glu-molecules than in the original simulation. This may be linked to the greater propensity for the 'circle'-motif to form between the C_α carboxyl and the amine in the original simulation, as the presence of a water molecule bridging the two groups may reduce the overall hydration as it is not close enough to the carboxyl to sit in the hydration shell, and inhibits other water molecules filling this gap.

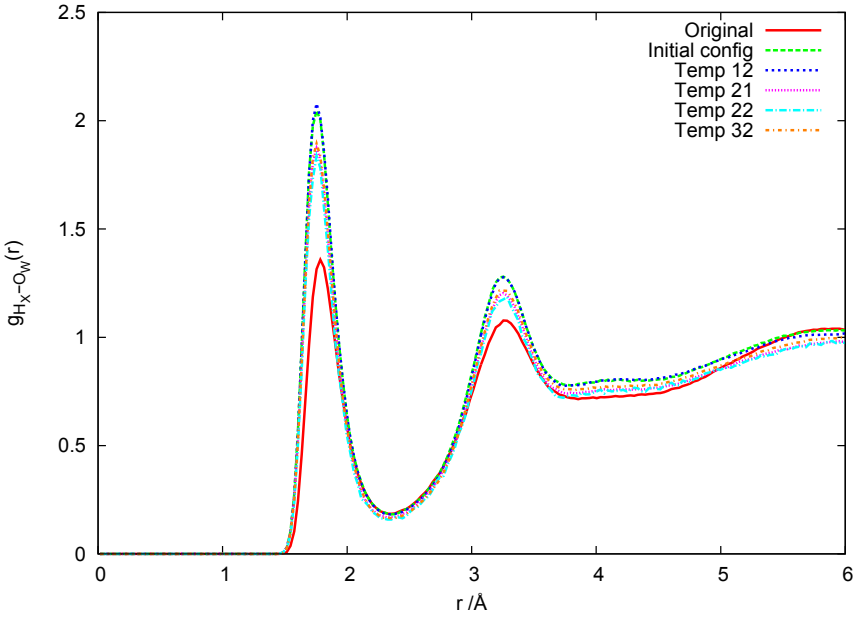
System	$n^N_{O_W}$
Original	2.43
Initial	2.83
Temp 12	2.82
Temp 21	2.58
Temp 22	2.60
Temp 32	2.70

Table 4.5: Co-ordination numbers n^β_α at a radial distance $r_{min} = 3.2$ for the glu-water $N - O_W$ interaction.

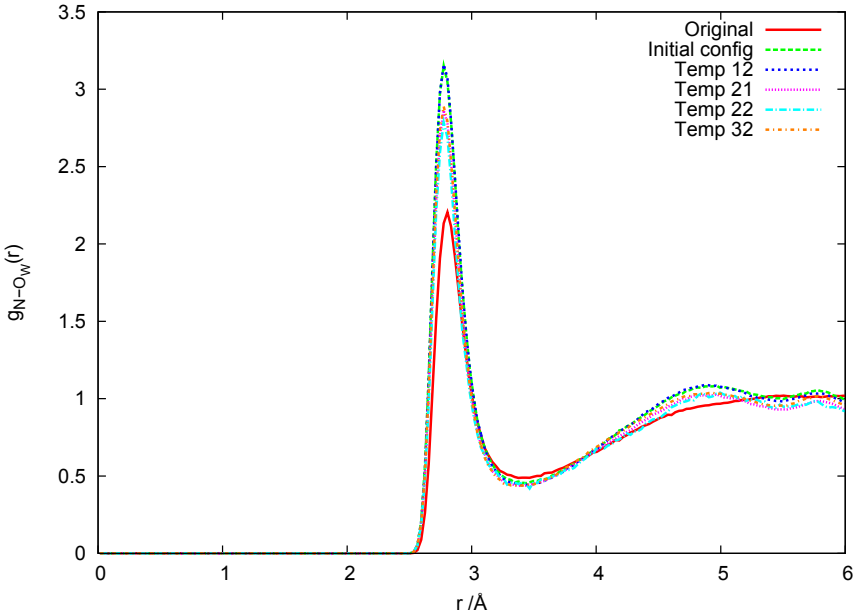
We next consider the 2D radial-radial distributions correlating the $r_{O_W - O_{CE}/O_{C1}}$ and $r_{O_W - O_{CE}/O_{C2}}$ distances, where O_{CE}/O_{C1} and O_{CE}/O_{C2} are the oxygen atoms within the same carboxyl group on the same glu-molecules (figures 4.24 and 4.25). Again we show the original system plots for the SPC/E system, along with the plots for the temperer 22 system.

For both the side-chain and the C_α-carboxyl distributions (figures 4.24 and 4.25 respectively), the plots are similar for the post-parallel-tempering simulations and the original simulation. The key difference observed in the original simulations between the $O_W - O_C$ and $O_W - O_{CE}$ plots is the greater propensity for the O_{CE} carboxyl-oxygens to form bifurcated bonds with water compared to the OC carboxyl-oxygens (shown by the greater intensity in the region around $_{O_W - O_{CE1}} \simeq r_{O_W - O_{CE2}}$). This is still true with the post-parallel-tempering simulations, although to a lesser degree with the $O_W - O_C$ plot, figure 4.25(b), showing a thin-band of higher intensity closer to the bifurcated region than the original simulation, figure 4.25(a). This is the case across all five post-parallel-tempering simulations (shown in appendix B, figures B.5 and B.6).

As the difference between the original and the new simulation 2D radial-

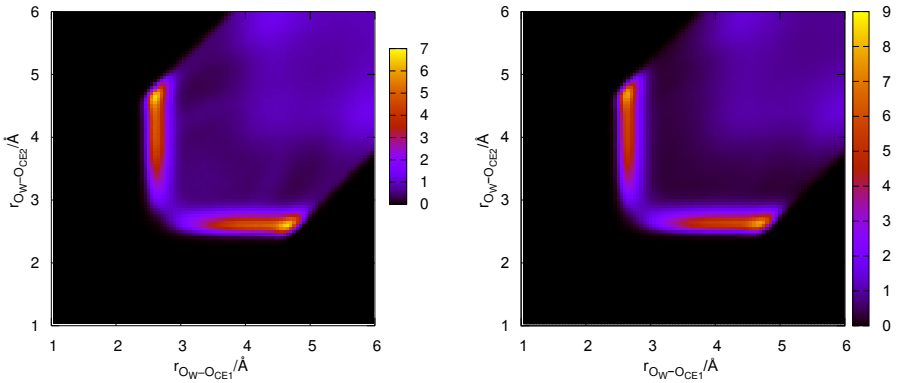


(a) HX-OW



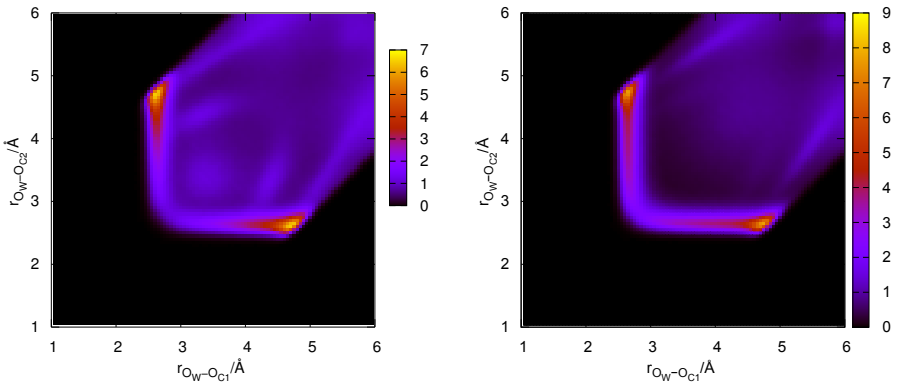
(b) NH3-OW

Figure 4.23: Radial distribution functions of the Glu-water amine-O_W interactions from the original simulations of CHARMM22 glu and SPC/E water discussed in chapter 3, and then the 5 systems simulated after parallel tempering was performed on the system.



(a) Original system $O_W - O_{CE}$ (b) Temperer 22 $O_W - O_{CE}$

Figure 4.24: Two-dimensional radial-radial distribution functions of the glu-water interactions between O_W (water) and O_{CE} (glu) atoms for the original system from chapter 3, and for the temperer 22 system. The plots for the other 4 systems are shown in appendix B, figure B.5.



(a) Original system $O_W - O_C$ (b) Temperer 22 $O_W - O_C$

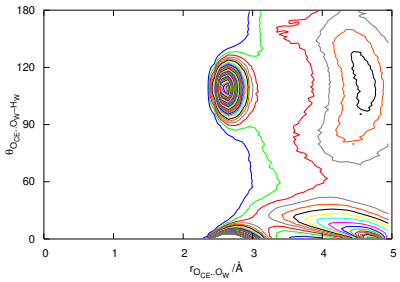
Figure 4.25: Two-dimensional radial-radial distribution functions of the glu-water interactions between O_W (water) and O_C (glu) atoms for the original system from chapter 3, and for the temperer 22 system. The plots for the other 4 systems are shown in appendix B, figure B.6.

radial plots between water and the carboxyl oxygens only occurs in the C_α interaction (figure 4.25(b)), and not the side chain interaction, it is possible that this enhanced prevalence of bifurcated bond states is related to the lack of the 'circle'-motif and therefore the difference in the $O_C - O_W$ RDFs in figure 4.21(b). If a bifurcated bond is formed, where one water molecule forms a bond with both carboxyl-oxygen atoms from the same carboxyl group, this reduces the ability of any other water molecules to form bonds with the same carboxyl. As the 'circle'-motif requires a 'stretched' single linear bond to form between an O_C atom and the water molecule, the presence of a bifurcated state may discourage the formation of the 'circle'-motifs.

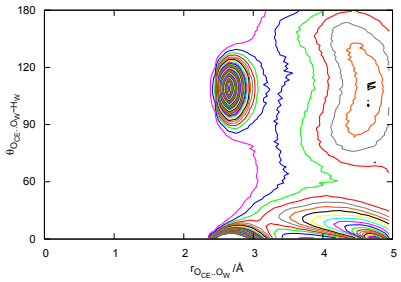
Figure 4.26 shows the 2D radial-angular distribution plots correlating the $O_{CE}/O_C - O_W$ radial distance with the $O_{CE}/O_C - O_W - H_W$ angle. Again the plots for the original and temperer 22 systems are shown, with the plots for the other four systems shown in appendix B (figures B.7 and B.8).

The plots correlating the $O_C - O_W$ radial distance and the $O_C - O_W - H_W$ angle for the original system and the temperer 22 system, figures 4.26(c) and 4.26(d) respectively, show similar motifs to each other, although there is a small difference between these plots and the corresponding O_{CE} plots (figures 4.26(a) and 4.26(b)). The O_C plots have a narrower peak in the region $r_{O_{CE}/O_C \cdot\cdot O_W - H_W} \simeq 4.8$ Å, $\theta_{O_{CE}/O_C \cdot\cdot O_W - H_W} \simeq 0°$ than the O_{CE} plots. This area of intensity is consistent with the co-ordination of the second, unbonded carboxyl-oxygen atom in the glu molecule co-ordinating with the O_W atom that is bonded to the first carboxyl-oxygen atom. The smaller angular and radial spread of the peak in the O_C plots compared to the O_{CE} plots suggests that the relative orientation of the water molecules around the C_α-carboxyl is more tightly constrained than around the side-chain carboxyl. This can be explained by the greater freedom of movement available to the side-chain, as it does not have the constraints of the amino-acid backbone.

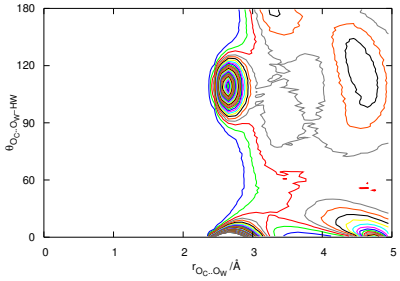
As the 2D radial-angular plots do not offer insight into the absence of the shoulder in the PINY_MD $O_C - O_W$ RDF we cannot at this point elucidate a reason for the difference. This problem is considered further in the discussion of the glutamate-sodium interactions.



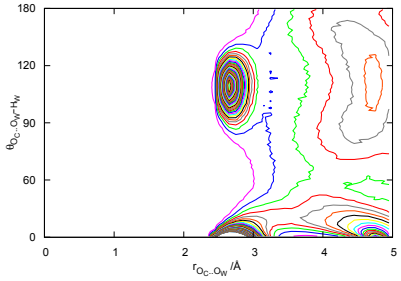
(a) Original system $O_{CE} \cdots O_W - H_W$



(b) Temperer 22 $O_{CE} \cdots O_W - H_W$



(c) Original system $O_C \cdots O_W - H_W$



(d) Temperer 22 $O_C \cdots O_W - H_W$

Figure 4.26: Two-dimensional radial-angular distribution functions of the glu-water interactions between O_{CE}/O_C atom and the water molecule, correlating the $O_{CE}/O_C - O_W$ radial distance with the $O_{CE}/O_C - O_W - H_W$ angle from the original simulation and the temperer 22 system. The plots for the other 4 systems are shown in appendix B, figures B.7 and B.8.

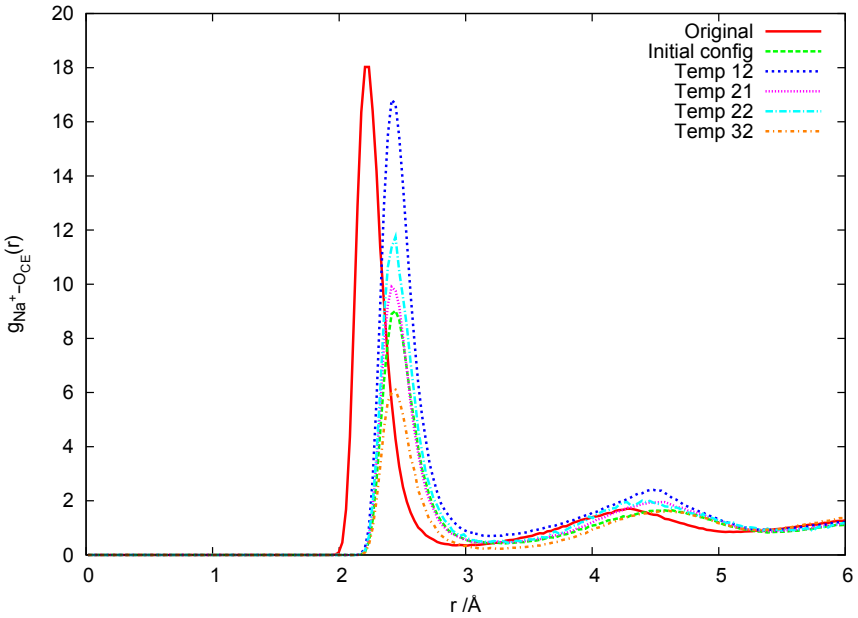
Glutamate-sodium Interactions

Figure 4.27 shows the glu-sodium RDFs for the $Na^+ - O_{CE}$ and the $Na^+ - O_C$ interactions (figures 4.27(a) and 4.27(a) respectively). The first key difference between the original and the post-parallel-tempering simulation data in these RDFs is the position of the first peak. The new simulations shift the first peak to larger r compared to the original simulation plot. This effect is usually due to the parametrisation of the force-fields used.

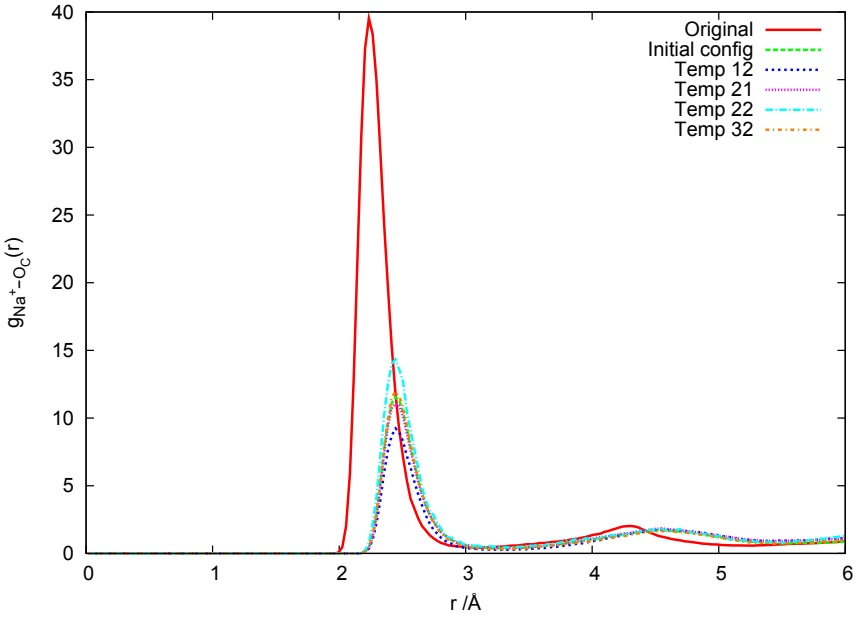
The original simulations, discussed in chapter 3 were performed in DLPOLY_2, using a parametrisation constructed from the CHARM22 forcefield [71] and downloaded from the CHARMM website [83]. The new simulations for this chapter were performed using PINY_MD, using the CHARMM22 potentials contained within the package. We have found that different versions of CHARMM22 provide different values for the sodium Lennard-Jones potential which governs inter-molecular interactions. Patra *et al.* noted that the original versions of CHARMM22 did not provide Lennard-Jones potential parameters for ions, but that CHARMM27 does [6]. Patra *et al.* also found that the version of CHARMM22 available from the XPlor website does contain ion parameters. In the construction of the glu force-field for the DLPOLY_2 simulations we have found that the current version of CHARMM22 available to download from the MacKerrell website [83] does contain Lennard-Jones parametrisations for ions. In the case of the sodium ion these values are the same in the CHARMM22 and CHARMM27 database. The PINY_MD simulation package, however, uses the XPlor-CHARMM22 parametrisations for sodium ions.

The effect of different potentials for the sodium ion is discussed in more detail in appendix D. In particular we summarise a study by Patra *et al.* which found that other key structural features of systems containing sodium ions parametrised with different Lennard-Jones potentials did not vary greatly [6]. This means that the basic thermodynamic and structural properties of two systems containing sodium ions with any of the Lennard-Jones potentials considered by Patra *et al.*, should still provide valuable physical information about such a system. This includes the two versions of the CHARMM22 potentials used in this work. In the following section we compare the original simulation interactions with the new post-parallel-tempering simulations.

Figure 4.27(b) shows that the intensity in the new system $Na^+ - O_C$ RDFs are



(a) $Na^+ - O_{CE}$



(b) $Na^+ - O_C$

Figure 4.27: Radial distribution functions of the glu-sodium interactions from the original simulations of CHARMM22 glu and SPC/E water discussed in chapter 3, and the post-parallel-tempering simulations.

all far smaller than in the original system. The $Na^+ - O_C$ occupations numbers, using a radial cut-off of 3.0 Å, are all closed to 0.5 (the original simulation, at the same cut-off has an coccupation number of 0.74), indicating that as the range of the first peak in the original system is at smaller r, that the change in occupation number is not as larger as it first appears on the RDF. This is not the case in the $Na^+ - O_{CE}$ RDFs, figure 4.27(a), where the original system is of similar intensity to the temperer 12 system RDFs, this is unlikely to be an artefact of the change in Lennard-Jones potential. It should be noted that Patra *et al.* did find a small increase in intensity from the X-Plor/CHARMM22 potential Na^+O_W RDFs to the CHARMM22/27 potentials [6]. As both the $Na^+ - O_{CE}$ and $Na^+ - O_C$ plots show a higher intensity for the original simulations than the new simulations this should be taken into consideration.

There is a spread of first-peak intensities in both the $Na^+ - O_{CE}$ and $Na^+ - O_C$ RDFs. We find that the temperer 12 system forms more $Na^+ - O_{CE}$ bonds than $Na^+ - O_C$ bonds, whereas the effect is reversed for the other post-parallel-tempering systems. In the original system the first-peak intensity in the $Na^+ - O_C$ RDF is more than double that of the $Na^+ - O_{CE}$ RDF, indicating a strong preference to associate sodium ions with the C_α-carboxyl. As this trend is reversed for the temperer 12 system, and significantly reduced for the other systems this suggests that this is due to the initial configuration of the system.

The original system is the only system to have such a marked difference in the first peak intensities of the $Na^+ - O_{CE}/O_C$ RDFs. As this is an artefact of the very large intensity in the $Na^+ - O_C$ RDF, this could be related to the shoulder in the $O_C - O_W$ RDF shown in figure 4.21(b), which is present in the original simulation but not in any of the new post-parallel-tempering simulations. On further analysis of the original simulation trajectory we were able to identify occasions where the sodium ion forms a bifurcated bond bridging both C_α-carboxyl atoms at the same time as a water molecule forms a 'circle'-motif with the same glu-molecule. A schematic of this is shown in figure 4.28. This is not the same motif as suggested by Daub *et al.* where a sodium ion mediates the carboxyl-water bond [45], which we were able to establish does not occur in the original simulation, as shown by the results discussed in chapter 3, as this does not involve a direct interaction between the water-molecule and the sodium ion. There is clearly some interaction occurring between the sodium ion and the water molecules, in that the water

molecules cannot approach the carboxyl-group closely while a sodium ion is in a bifurcated motif with the carboxyl-oxygens. It is reasonable to assume that the water molecules' proximity to the amine group, at the same time as the carboxyl-group, is a side effect of the water molecule approaching a carboxyl group that is currently bonded to a sodium ion. The water molecule's position is presumably stabilised by the presence of the amine, while still attracted by the negative charge of the carboxyl group. Although the O_W atom is repelled by the sodium ion, the water molecule can remain in a stable configuration in the 'circular' motif described in chapter 3, and thereby producing the shoulder in the original-system $O_C - O_W$ RDF.

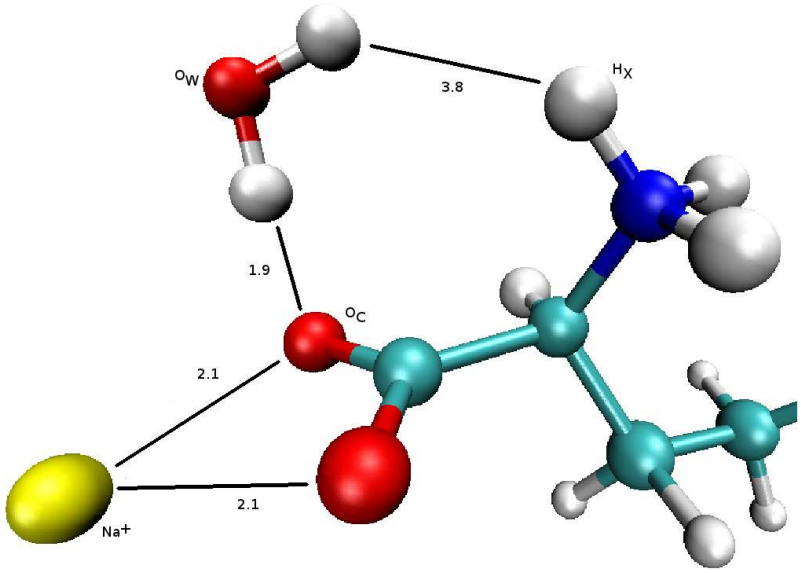
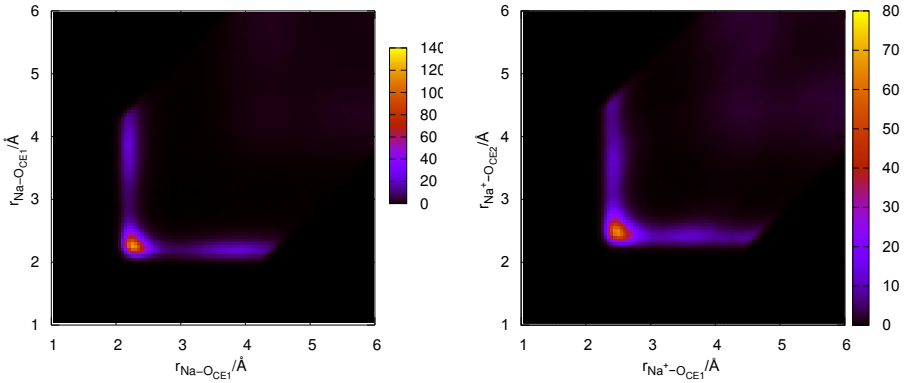


Figure 4.28: Schematic of glu-water circle motif in the original system where a water molecule bonds to both the C_α-carboxyl and the amine group of the same glu-molecule. Here we find that this when a sodium ion forms a bifurcated bond with the same C_α-carboxyl-oxygens.

As the number of $Na^+ - O_C$ bonds is smaller, relative to the $Na^+ - O_{CE}$ bonds, in the PINY_MD simulations, this therefore also helps to explain the smaller frequency of circular motifs in the post-parallel-tempering simulations, and therefore the absence of the shoulder in the $O_C - O_W$ RDF.

To verify the presence of sodium-carboxyl bifurcated bonds, figures 4.29 and 4.30 show the 2D radial-radial distribution plot correlating the $r_{Na^+-O_{CE}/O_{C1}}$ and $r_{Na^+-O_{CE}/O_{C2}}$ radial distances. The plots all have a high intensity peak in the region $r_{Na^+-O_{CE}/O_{C1}} \simeq r_{Na^+-O_{CE}/O_{C2}} \simeq 2.4$ Å, indicating a preference to form

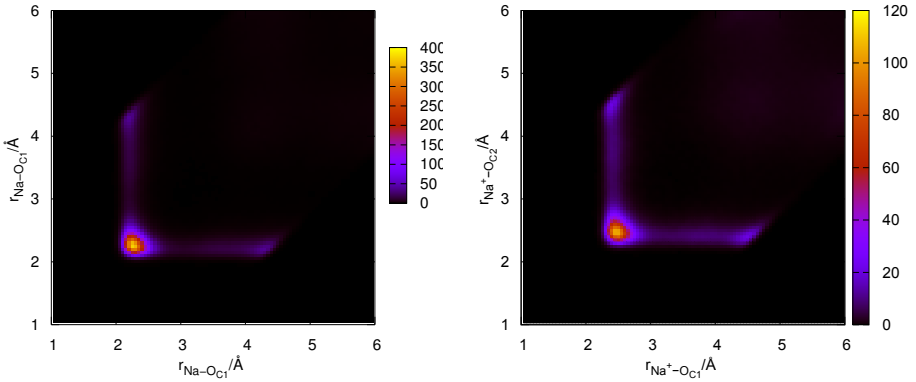
bifurcated motifs where the sodium ion correlates with both carboxyl-oxygens. Similar plots were obtained for the other four systems and are shown in figures B.9 and B.10.



(a) Original system $Na^+ - O_{CE}$ (b) Temperer 22 $Na^+ - O_{CE}$

Figure 4.29: Two-dimensional radial-radial distribution functions of the sodium-glu interactions between Na^+ (sodium ion) and O_{CE} (glu) atoms, correlating the $r_{Na^+-O_{CE1}}$ and $r_{Na^+-O_{CE2}}$ radial distances,, for the original system and the temperer 22 system.

4.2.5 Water-Sodium Interactions



(a) Original system $Na^+ - O_C$ (b) Temperer 22 $Na^+ - O_C$

Figure 4.30: Two-dimensional radial-radial distribution functions of the sodium-glu interactions between Na^+ (sodium ion) and O_C (glu) atoms, correlating the $r_{Na^+-O_{C1}}$ and $r_{Na^+-O_{C2}}$ radial distances, for the original system and the temperer 22 system.

Figure 4.31 shows the RDF corresponding to the $O_W - Na^+$ interaction. Again we see that the new simulations all have a first-peak at larger r than

the original simulation. There are no other key differences between the original and post-parallel-tempering simulations, suggesting that the properties of the $O_W - Na^+$ interaction are not greatly affected by either the change in Lennard-Jones potential or the increased sampling of the parallel tempering.

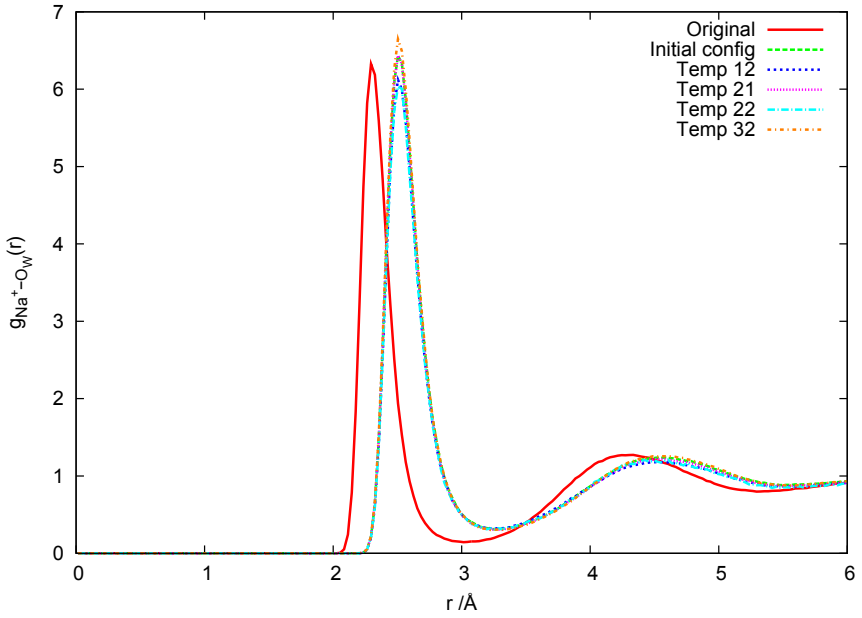
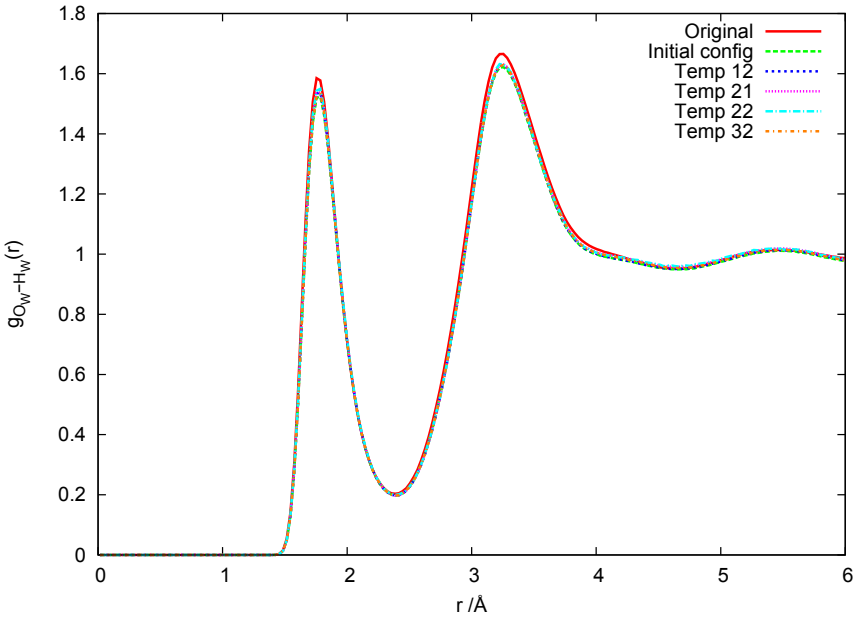


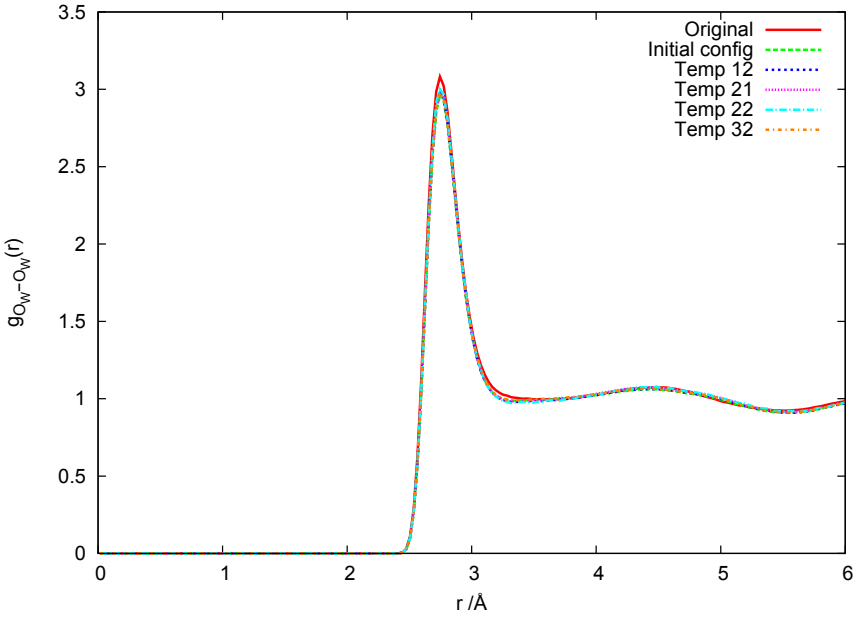
Figure 4.31: Radial distribution functions of the water-sodium ($O_W - Na^+$) interactions from the original simulations of CHARMM22 glu and SPC/E water discussed in chapter 3, and then the 5 systems simulated after parallel tempering was performed on the system.

Water-water interactions

For completeness we show the water-water RDFs in figure 4.32. As with the sodium-water RDFs the new the parallel tempering has had only a negligible effect on the water within the glu solutions. This indicates that the change in glu cluster sizes in the different post-parallel-tempering simulations has not had a significant impact on the bulk water in the systems relative to the original simulation or the initial configuration system.



(a) $O_W - H_W$



(b) $O_W - O_W$

Figure 4.32: Radial distribution functions of the water-water interactions from the original simulations of CHARMM22 glu and SPC/E water discussed in chapter 3, and then the 5 systems simulated after parallel tempering was performed on the system.

4.2.6 Conclusions from parallel tempering simulations of aqueous glutamate

Using the parallel tempering technique with Molecular Dynamics simulations has enabled better sampling of the possible conformations of the aqueous glutamate (glu) solution. Comparing this with our classical simulations discussed in chapter 3 has highlighted the limited sampling that the original classical simulation had of the system configurations. Many of the structural features in the original simulation are still present in the post-parallel-tempering simulations.

We focused on obtaining different glu cluster size distributions from the parallel tempering simulations. These were then subjected to a 2 ns classical simulation before examining the different structures obtained. The key difference, other than cluster sizes, was a difference in the preference for glu-glu bonding sites. In the initial simulations we found that glu-glu bonds preferred to form between the amine and the side-chain carboxyl-group, not the C_α-carboxyl. Of the five new simulations considered, only one preferred to form glu-glu bonds*via* the C_α-carboxyl, indicating that a slight preference for side-chain glu-glu bonds still exists, although initial configuration of the system plays a large part in the preference. We also found that the starting configuration of the system affected the propensity for bifurcated bonds to form on each of the carboxyl-sites, although both carboxyl groups still formed bifurcated bonds with the amine group of adjacent glu-molecules, though to a much lesser extent.

In the post-parallel-tempering simulations we also found that PINY_MD uses a different version of the CHARMM22 parameters (the X-Plor/CHARMM22 potential) than those used in the DLPOLY_2 simulations (the CHARMM22/27 potential). This results in the Lennard-Jones potential of the sodium ion interactions with other atoms being slightly different, although the majority of the structural properties of the system are unaffected (see appendix D for more details). Aside from the difference in the bond lengths of the sodium-glu and sodium-water interactions, one key effect of the different potential was the removal of the shoulder on the $O_C - O_W$ RDF. This shoulder, between the first and second peaks in the RDF, was present in the original DLPOLY_2 simulations using the CHARMM22/27 potential, however, it is not present in any of the PINY_MD X-Plor/CHARMM22 simulations. We had previously identified that the shoulder was due to a 'circle'-motif, where a water molecule bonds to the amine and the

C_α carboxyl group of the same glu-molecule. This stretches the water-O_C bond-length, but maintains its stability, resulting in the $O_C - O_W$ RDF shoulder. In the new, post-parallel-tempering simulations, this circular motif still exists, but in far fewer cases, resulting in the removal of the shoulder in the RDF.

The $O_{CE} - O_W$ occupation numbers are consistently greater than the $O_C - O_W$ for all five systems, indicating a preference for glu-water bonds to form *via* the side chain. In a biological context when glutamate is combined with other amino acids to form a polypeptide, it would be the side-chain that would form bonds with water and other species. As a single glutamate molecule prefers to bond with water *via* the side-chain, the freedom this produces for the C_α-carboxyl may make the condensation reaction to forming the peptide bond easier to achieve as it may be able to associate with other amino acids more easily.

We identified that the presence of the sodium ion, bonding to the C_α-carboxyl, influenced the ability of the system to form 'circle'-motifs. We suggest that this is because the sodium, being positively charged, inhibits the approach of the H_W atom of the water molecule. The water molecule is still attracted to the negative charge of the carboxyl-oxygens, and maintains a stable position due to the presence of the amine group which it is currently co-ordinated with. In the new simulations, firstly there are fewer sodium-O_C interactions, providing fewer opportunities for this interaction to occur, and secondly, the radial distance of the sodium-O_C bond is slightly greater in the PINY_MD simulations due to the different Lennard-Jones potential. We suggest that this increase in the sodium-O_C radial distance enables the closer approach of water molecules to the carboxyl group. At this point the water molecule no longer interacts with the amine group, removing the stability it obtained from that interaction. This results in far fewer 'circle'-motifs being formed and the removal of the corresponding shoulder in the $O_C - O_W$ RDFs.

If the circle motif exists as a biological phenomenon, which cannot be ruled out despite the variation with the sodium ion potential used, the its existence may change the behaviour of glutamate in the body. Stabilising a water molecule in this way, making it a relatively long-lived contact that is dependent on the presence of sodium ions, may extend to the presence of other ions, to a greater or lesser extent. As the water bonds to the backbone of the molecule this may inhibit the condensation reaction to form peptide bonds and ultimately prohibit

protein formation, or reduce the reaction speed. This may provide a mechanism for controlling the speed of the reaction. The circle motifs dependence on the presence of the sodium ion may also effect the transport of glutamate molecules across membranes by sodium ions. If this effect does exist in a biological system for sodium, but not for other ions, then it other ions may provide an alternative transport mechanism for artificially increasing the rate of transfer.

Studies of other amino acids in solution with sodium ion may also provide examples of the circle motif, as the effect relies on the existence of an ionised backbone of amino acids. To our knowledge no studies have identified this motif. As the only other negatively charged, zwitterionic, amino acid, aspartic acid would provide a useful comparison as it may also have been studied in solution with sodium ions. To our knowledge no Molecular Dynamics simulations exist of aspartic acid in aqueous solution with sodium ions.

These results highlight the importance of enhanced sampling in Molecular Dynamics due to the limitations of the short time-scales of such simulations. This is particularly important in the case of systems with slow dynamics such as the glu system considered here. This may be particularly relevant to biomolecular systems which often have a complex energy landscape, inhibiting the transition between configurations. Our method of setting up the simulation and comparing results could be applied to a wide variety of systems, though of course the specific details regarding the mechanics of the simulation (temperer spacing and upper limits etc.) will still have to be devised on a case-by-case basis.

Finally, it is important to consider side-effects of the empirical potentials used. As discussed here even small changes may produce interesting results. These are not necessarily incorrect but should always be considered as a possible artefact of the potential. In particular, these results suggest that, at least for this instance, the commonly held belief that CHARMM22 potential produces over-bifurcated interactions may not be necessarily true, and we would emphasise the importance of comparison with other data, particularly with first-principles simulations.

Chapter 5

The structure and dynamics of the GPE tripeptide

The chapter considers the structure of the tripeptide L-glycine-L-proline-L-glutamate (gly-pro-glu or GPE), and is split into three sections. Firstly, we provide a summary of the simulation details and the set-up procedures used. Secondly, we consider four systems containing a single GPE molecule in a box of water, each using a different conformational isomer of the GPE molecule. Finally, we consider three initial conformations of the GPE molecule, each simulated in solution with water and sodium counter-ions at two different concentrations, resulting in a total of six systems to be considered for analysis.

In the following analysis we follow a labelling convention similar to that used in the glutamate simulations. We refer to the atoms of water molecules as O_W and H_W for the oxygen and hydrogen atoms respectively. Sodium ions are labelled Na^+ and the labelling of the GPE atoms is shown in figure 5.1.

5.1 Simulation Details

The process of obtaining a GPE motif for simulation in water, discussed in detail below, resulted in the simulation of various conformations of the GPE molecule being studied. A summary of the development of the GPE simulations is shown in the chart in figure 5.2. This shows the progression from the initial structures obtained from the protein databank [2], through to the final six simulations of GPE molecules carried out at different concentrations. This process is discussed

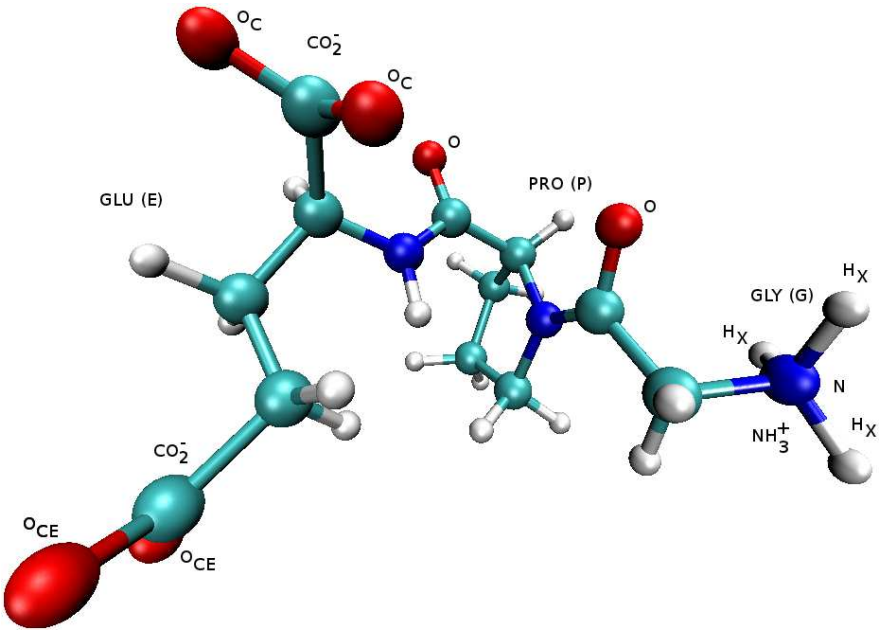


Figure 5.1: Labelling scheme for the GPE molecule.

in more detail in the rest of the chapter, however a summary of all the systems studied is given in table 5.1 for reference.

All simulations of the GPE tripeptide carried out for this work were performed using the PINY_MD simulation package [109] with the CHARMM22 classical potential [71] for the parametrisation of GPE and sodium ions, and the SPC/E water potential [82] for water molecules. The SPC/E water potential was used as the classical glutamate simulations studied in chapter 3 showed similar results for the SPC/E and F3C parameterised water molecules. SPC/E is a simpler model, requiring fewer computational resources than F3C, and therefore this was chosen over the F3C potential for the simulations of aqueous GPE. CHARMM22 was originally designed to be used with the TIP3P water potential. However, Mark et al. [73], show that the TIP3P potential does not reproduce the tetrahedral structure of water accurately, and therefore we have not used it in this study.

All Molecular Dynamics simulations were simulated using 1 fs time-steps, with a Nosé-Hoover chain length of 2. The Energy Minimisation procedure uses the conjugate gradient method described in Chapter 2.

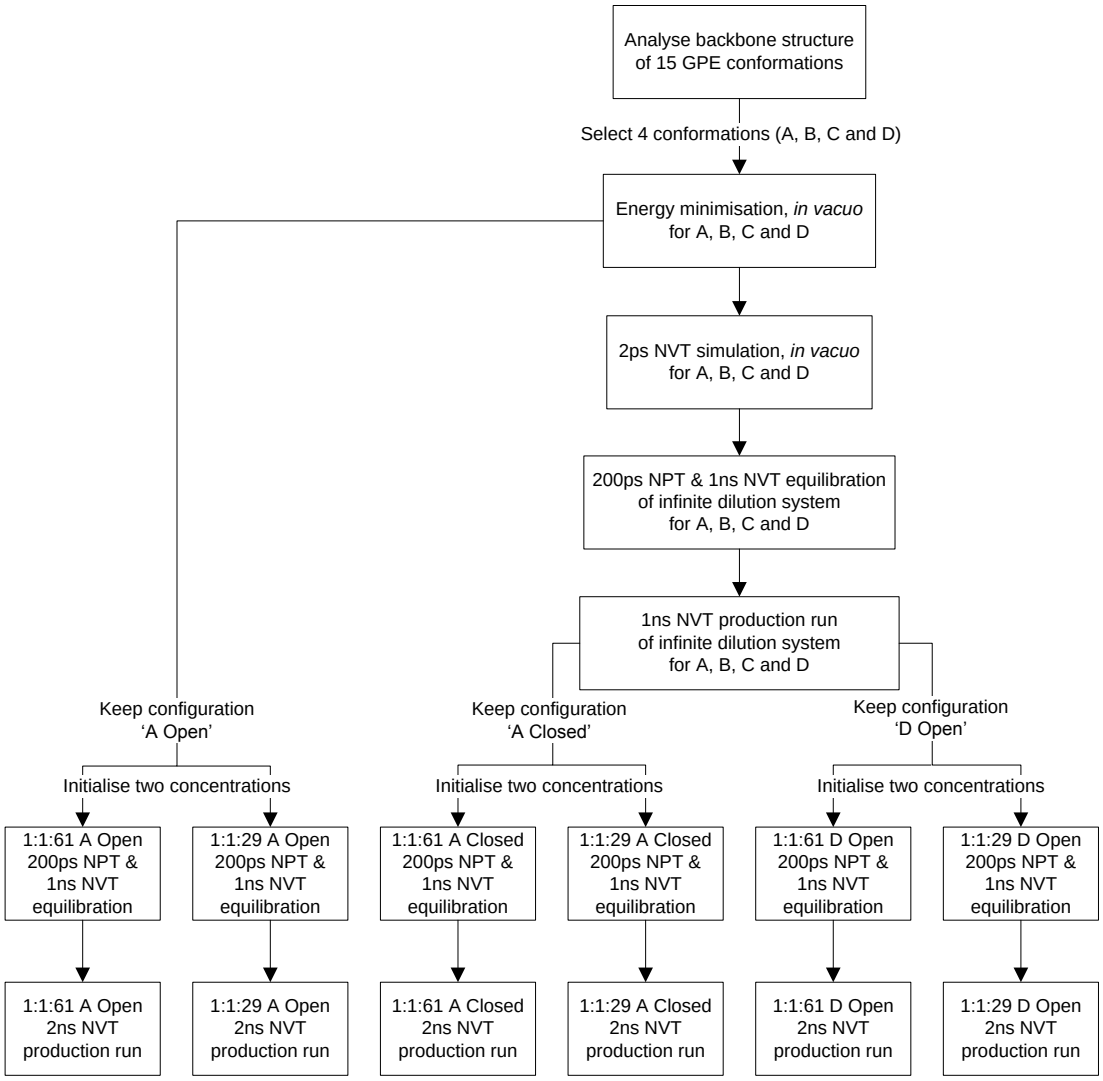


Figure 5.2: Chart of different simulations carried out on the GPE molecule.

Simulation	Length	Thermostat
Initialisation; in *vacuo* with sodium		
Structure A Energy Min		N/A
Structure B Energy Min		N/A
Structure C Energy Min		N/A
Structure D Energy Min		N/A
In *vacuo* simulations		
Structure A *in vacuo* MD	2 ps	NVT
Structure B *in vacuo* MD	2 ps	NVT
Structure C *in vacuo* MD	2 ps	NVT
Structure D *in vacuo* MD	2 ps	NVT
Single GPE in water Pressurisation EQM		
Structure A MD	200 ps	NPT
Structure B MD	200 ps	NPT
Structure C MD	200 ps	NPT
Structure D MD	200 ps	NPT
Single GPE in water EQM		
Structure A MD	1 ns	NVT
Structure B MD	1 ns	NVT
Structure C MD	1 ns	NVT
Structure D MD	1 ns	NVT
Single GPE in water production run		
Structure A Production MD	1 ns	NVT
Structure B Production MD	1 ns	NVT
Structure C Production MD	1 ns	NVT
Structure D Production MD	1 ns	NVT
1:1:29 GPE:Sodium:Water Simulations		
A closed EQM MD	200 ps	NPT
A open EQM MD	200 ps	NPT
D open EQM MD	200 ps	NPT
A closed EQM MD	1 ns	NVT
A open EQM MD	1 ns	NVT
D open EQM MD	1 ns	NVT
A closed Production MD	2 ns	NVT
A open Production MD	2 ns	NVT
D open Production MD	2 ns	NVT
1:1:61 GPE:Sodium:Water Simulations		
A closed EQM MD	200 ps	NPT
A open EQM MD	200 ps	NPT
D open EQM MD	200 ps	NPT
A closed EQM MD	1 ns	NVT
A open EQM MD	1 ns	NVT
D open EQM MD	1 ns	NVT
A closed Production MD	2 ns	NVT
A open Production MD	2 ns	NVT
D open Production MD	2 ns	NVT

Table 5.1: Simulations carried out on the GPE molecule. EQM refers to an equilbration procedure.

5.2 Set-up of the GPE system

The initial configuration of the GPE molecule was obtained from the protein-databank, as a fragment of the 55-72 segment of staphylococcal nuclease [2, 5]. The GPE fragment is at the end of the 55-58 segment of the L optical isomer polypeptide such that the glycine (G) residue is capped with the N-terminal (NH_3^+). We then capped the glutamate (E) residue with a C-terminal, CO_2^-, and the carboxyl of the glutamate residue was de-protonated to produce a second CO_2^-. This results in the GPE configuration seen in figure 5.1.

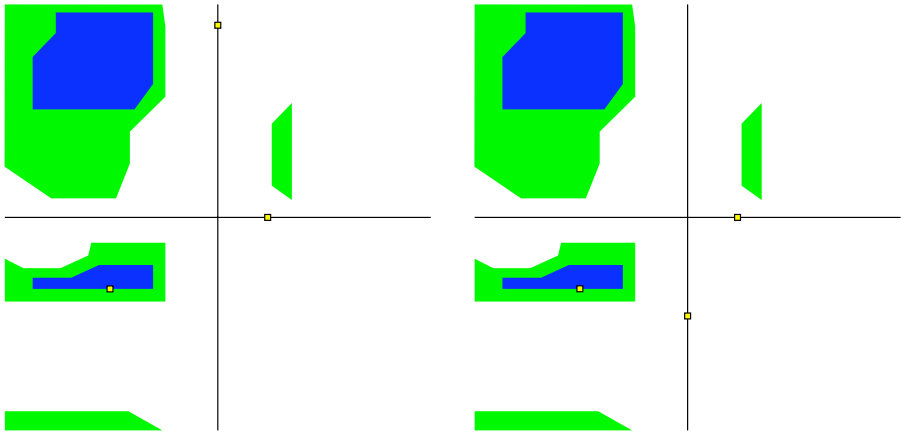
The GPE molecule has three charged sites, an amine, NH_3^+, and two carboxyls, CO_2^-, resulting in an overall charge of -1 eV, as we had in the glutamate molecule considered in chapters 3 and 4.

Fifteen unique GPE structures, all obtained by two-dimensional (2D) nuclear magnetic resonance (NMR) of the staphylococcal nuclease in water, are available from the protein databank, provided by Wang *et al.* in their study of the conformational freedom this nuclease. All fifteen fragments were downloaded and capped. Tests were then performed on all fifteen GPE structures to assess the conformational uniqueness of the tripeptide molecule.

5.2.1 Analysis of the different GPE conformations

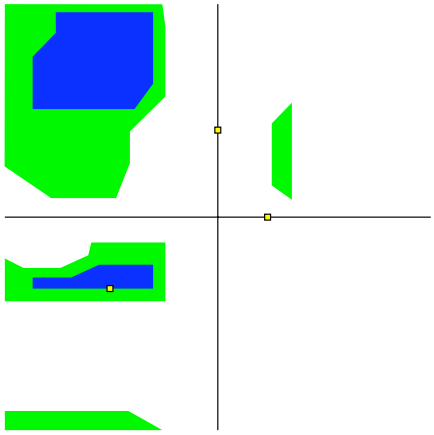
Each of the 15 configurations obtained from the protein databank had the potential to be conformationally unique. If the conformations are sufficiently different then transition between them may not be possible in the time-scales of traditional MD simulations. Thus we consider all fifteen motifs. To assess which structures were truly individual we considered the orientation of the backbone of the peptide through the use of Ramachandran maps of the fifteen structures (see figures 5.3, 5.4 and 5.5).

Each Ramachandran map has three points; one for each of the backbone residues, glycine (gly), proline (pro) and glutamate (glu). The gly orientational point lies on the ψ-axis of the map; at this point the ϕ rotational angle is zero as it is a terminal point of the polypeptide backbone. Similarly, the point corresponding to the glu residue lies on the ϕ-axis as it is the other terminal. The point in the bottom left hand quadrant (the α-helix region) is due to the angular correlation of the ϕ-ψ orientations for the pro residue.

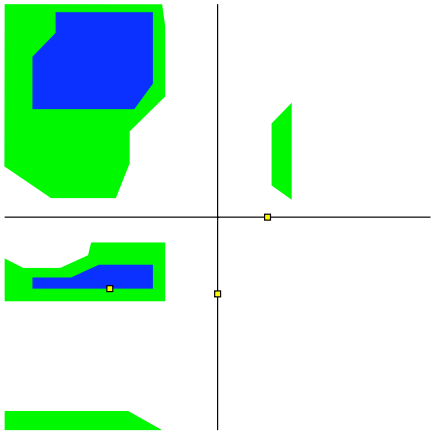


(a) Structure 1

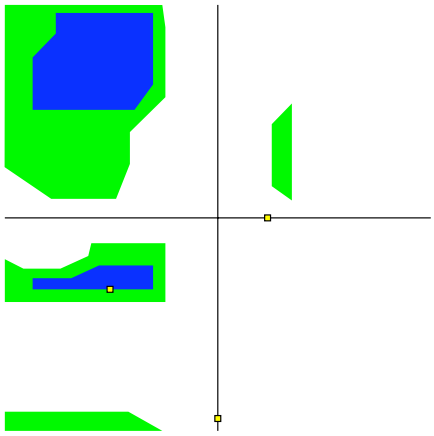
(b) Structure 2



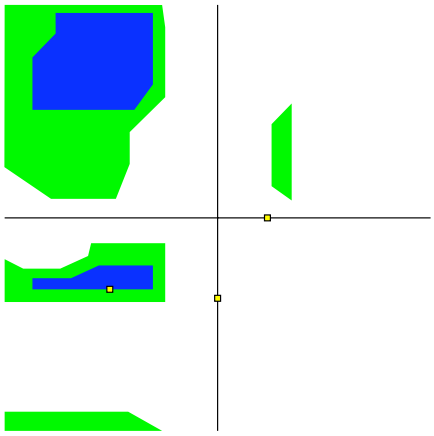
(c) Structure 3



(d) Structure 4

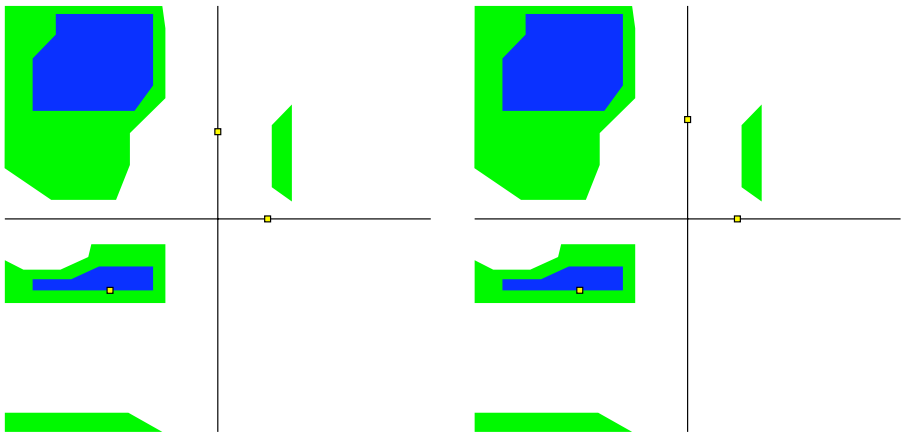


(e) Structure 5



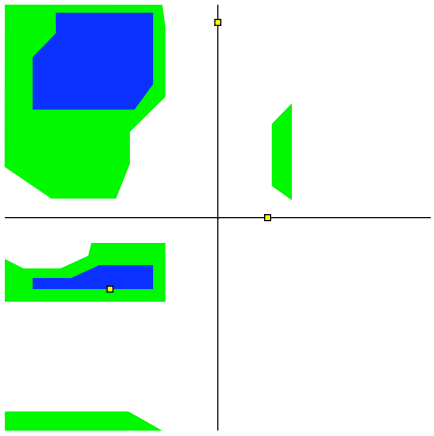
(f) Structure 6

Figure 5.3: Ramachandran maps for the three residues of the first six GPE conformations, as obtained from protein databank files of the 55-72 segment of staphylococcal nuclease [5]. Plots are of the torsional angles of the peptide backbone, ϕ (x-axis) against ψ (y-axis).

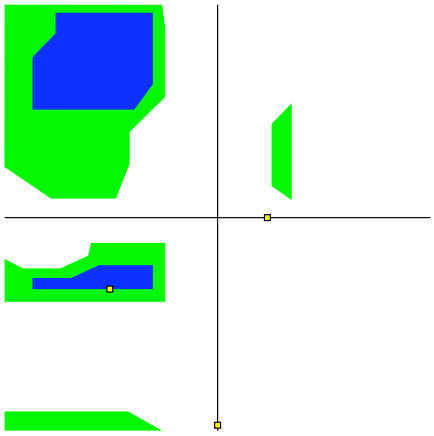


(a) Structure 7

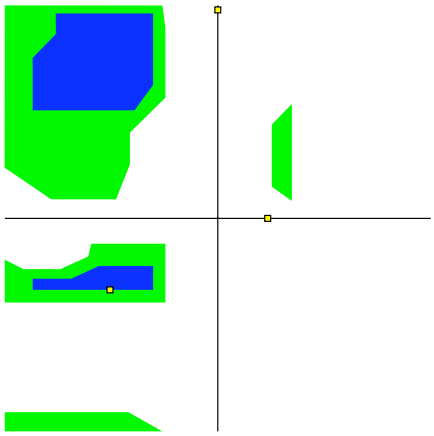
(b) Structure 8



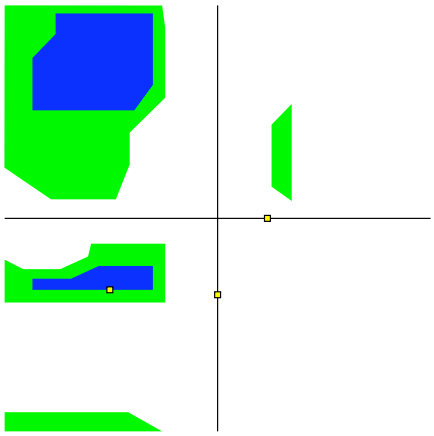
(c) Structure 9



(d) Structure 10

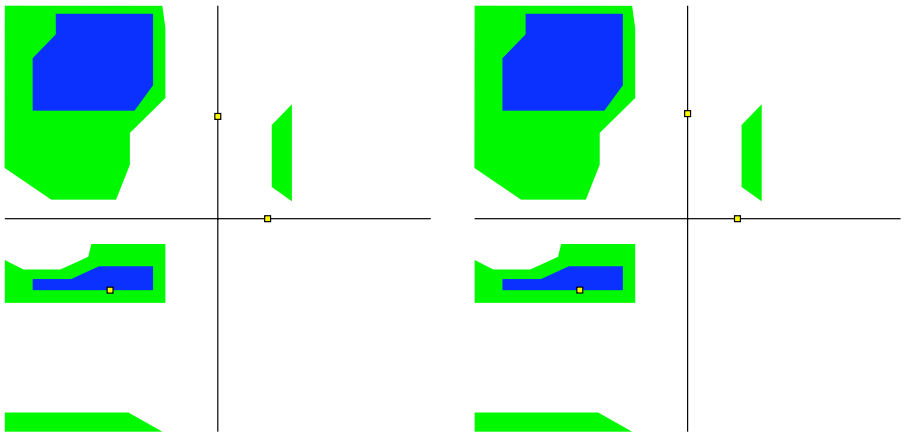


(e) Structure 11



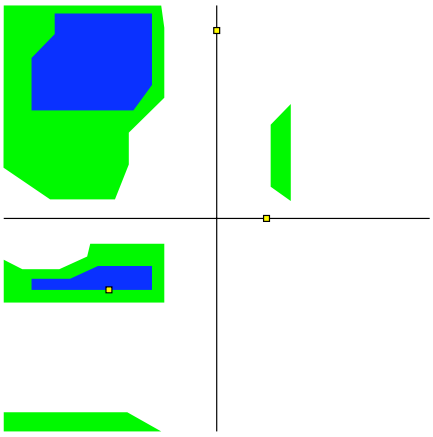
(f) Structure 12

Figure 5.4: Ramachandran maps for the three residues for GPE conformations 6-12, as obtained from protein databank files of the 55-72 segment of staphylococcal nuclease [5]. Plots are of the torsional angles of the peptide backbone, ϕ (x-axis) against ψ (y-axis).



(a) Structure 13

(b) Structure 14



(c) Structure 15

Figure 5.5: Ramachandran maps for the three residues of the last three GPE conformations, as obtained from protein databank files of the 55-72 segment of staphylococcal nuclease [5]. Plots are of the torsional angles of the peptide backbone, ϕ (x-axis) against ψ (y-axis).

The position of the pro-angle torsional angle point is similarly placed in all fifteen Ramachandran maps, as are the positions of the glu-angles. The key differences in the conformation of the backbone of the tripeptide originate from the orientation of the gly residue. The position of this point is in one of four distinct regions in the map. These are summarised in table 5.2.

Group	Structures	Orientation of gly-residue	
		ϕ	ψ
A	1, 9, 11, 15	0°	162°
B	2, 4, 6, 12	0°	-83°
C	3, 7, 8, 13, 14	0°	73°
D	5, 10	0°	169°

Table 5.2: Position of gly-residue angles on the Ramachandran plot for the fifteen different initial configurations.

We find that the fifteen GPE configurations all have one of four distinct orientations of the gly-residue, and hence we conclude that there are four principle conformations to be further investigated. To further assess the uniqueness of these four individual structures, hereafter referred to as A, B, C and D (using initial configurations of structures 1, 2, 3 and 5 respectively), we performed an energy minimisation and then a preliminary classical Molecular Dynamics (MD) simulation of the molecule in *vacuo* with a sodium counter-ion.

For the energy minimization, or configuration optimisation, we used the classical force-field based conjugate gradient method. This iteratively optimises the geometry of the atoms until the force convergence measure, $\Delta F = [\frac{1}{N}\sum_{i=1}^{N} F_i.F_j]$, is less than a threshold, where N refers the number of atoms in the system. Using a threshold of F<0.001 kJ mol^{-1}Å^{-1}, this method provides an optimal configuration, subject to the constraints of the classical force-field, of the tripeptide from the initial conformation obtained from the protein databank.

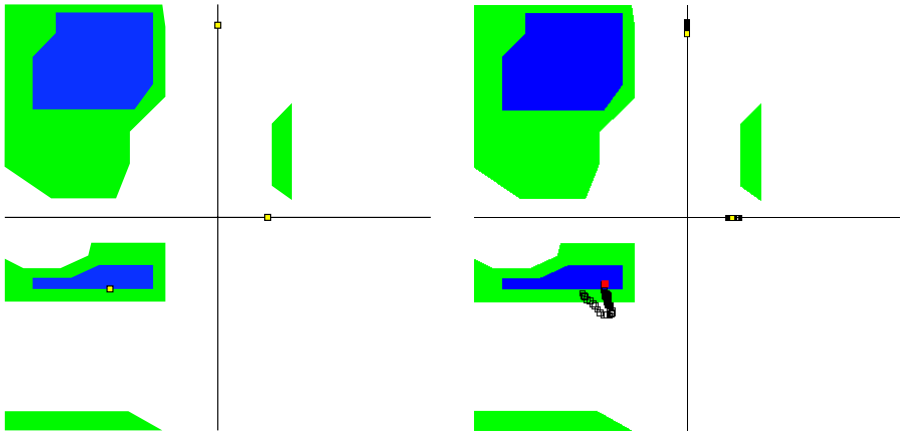
The total number of energy minimisation, conjugate gradient steps required for each of the four unique structures to reach the force threshold is shown in table 5.3. From this we see that structures A and D were much further away from an equilibrium structure than structures B and C. Once all four structures reached the threshold the molecular conformations of each structure should be similarly stable with respect to the CHARMM22 force-field.

Structure	Number of CG steps
Structure A	1147
Structure B	100
Structure C	494
Structure D	1054

Table 5.3: Number of steps of the conjugate gradient algorithm required for the system to converge to a threshold force of $F < 0.001$ kJ mol^{-1}Å^{-1}.

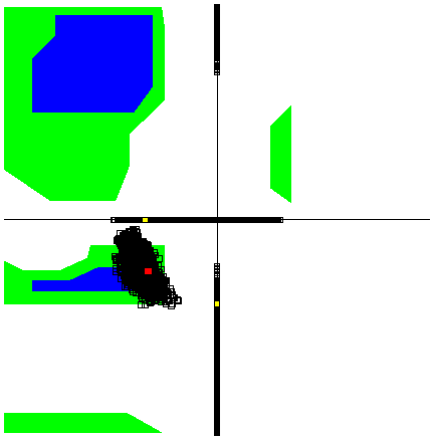
After the energy minimisation, the GPE and sodium ions were subjected to a further in *vacuo* MD simulation using the NVT Nosé-Hoover thermostat, at 300 K, for 2000 timesteps of 1 fs each, producing a simulation trajectory of 2 ps. The trajectories obtained for both the energy minimisation steps and the in *vacuo* MD simulation were analysed using Visual Molecular Dynamics (VMD) software [125] to produce Ramachandran maps for the 'backbone' angles of the tripeptide. The Ramachandran maps of the initial structure, the energy minimisation steps and the MD steps for each of the four systems are shown in figures 5.7 and 5.7. The energy-minimisation and the initial in *vacuo* MD simulation show a time-resolved plot where the torsional angles are plotted for all steps in the simulation.

Although each of the four systems starts with a distinctly different gly torsional orientation, the Ramachandran maps show that during the initial 2 ps MD simulation each of the four systems covers a wide region of the ψ-axis (y-axis). The final position of the yellow dot on the y-axes $\phi = 0°$ of the plots shows the angle at approximately the same orientation for all four structures at the end of the simulation. The pro angle (red dot in the bottom-left quadrant) is significantly different at the end of the MD simulation for structure C than the other three molecules. Covering this breadth of angular orientations in such a short simulation time suggests little energetic impediment to the angle of the gly peptide bond. As these simulations are performed in *vacuo*, and because there is also a large variation in the pro-torsional angles, all four systems are studied further in solution with water.

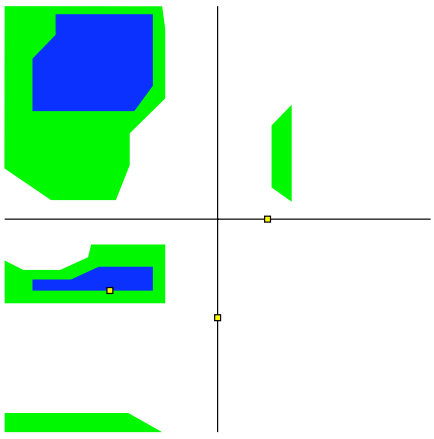


(a) Structure A Initial

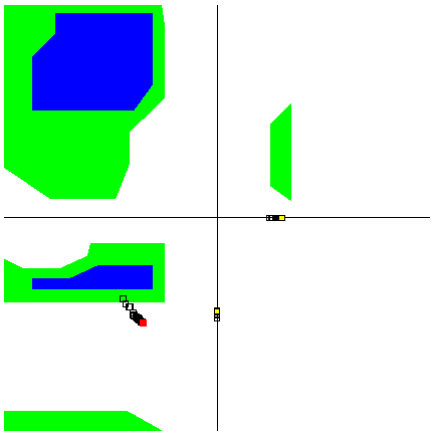
(b) Structure A Energy Min



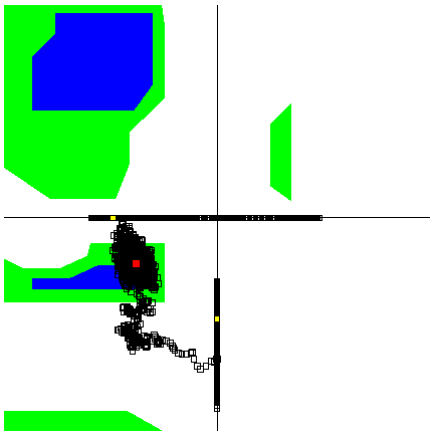
(c) Structure A MD Sim



(d) Structure B Initial

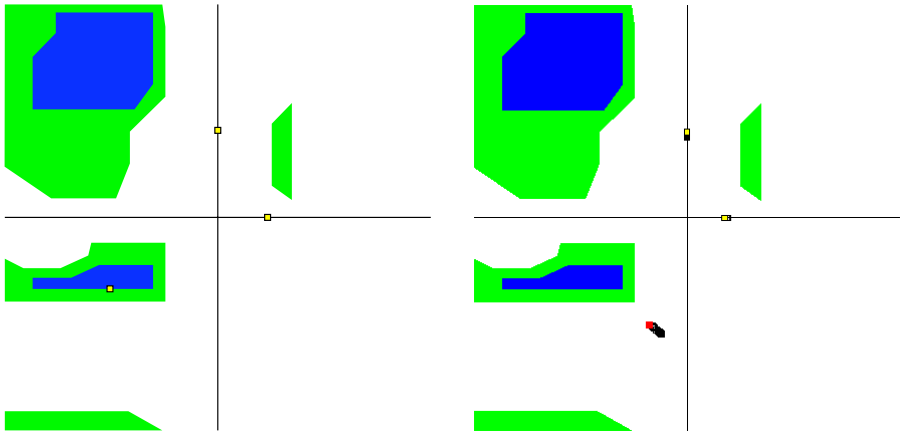


(e) Structure B Energy Min



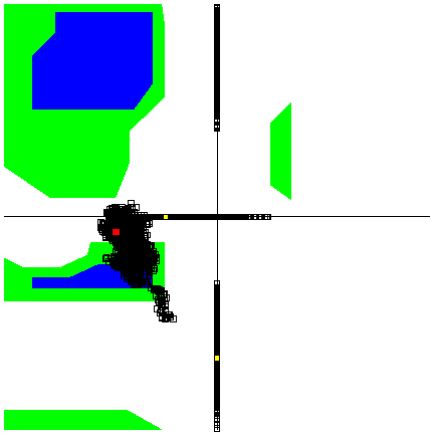
(f) Structure B MD Sim

Figure 5.6: Ramachandran maps for GPE structures A and B, showing the maps for the initial structure, the structure during the energy minimisation and during the MD simulation. Maps plot the torsional angles of the peptide bonds ϕ (x-axis) against ψ (y-axis).

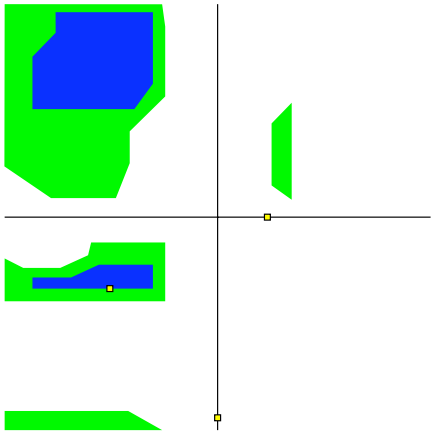


(a) Structure C Initial

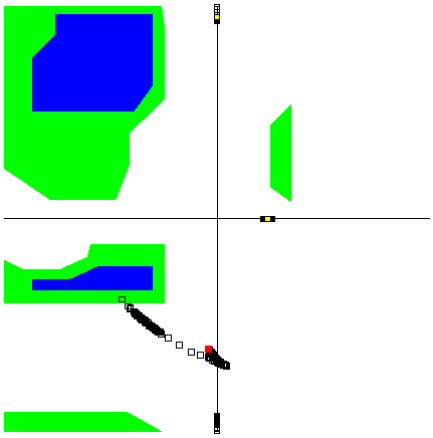
(b) Structure C Energy Min



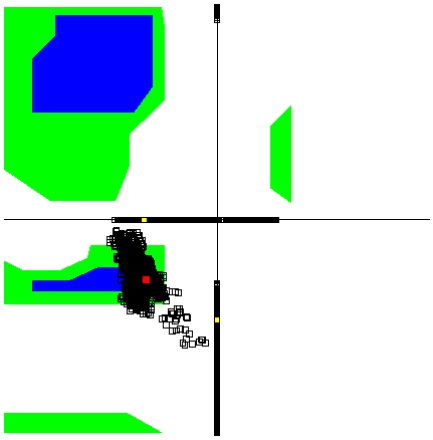
(c) Structure C MD Sim



(d) Structure D Initial



(e) Structure D Energy Min



(f) Structure D MD Sim

Figure 5.7: Ramachandran maps for GPE structures C and D, showing the maps for the initial structure, the structure during the energy minimisation and during the MD simulation. Maps plot the torsional angles of the peptide bonds ϕ (x-axis) against ψ (y-axis).

5.3 Simulations of a single GPE molecule in water

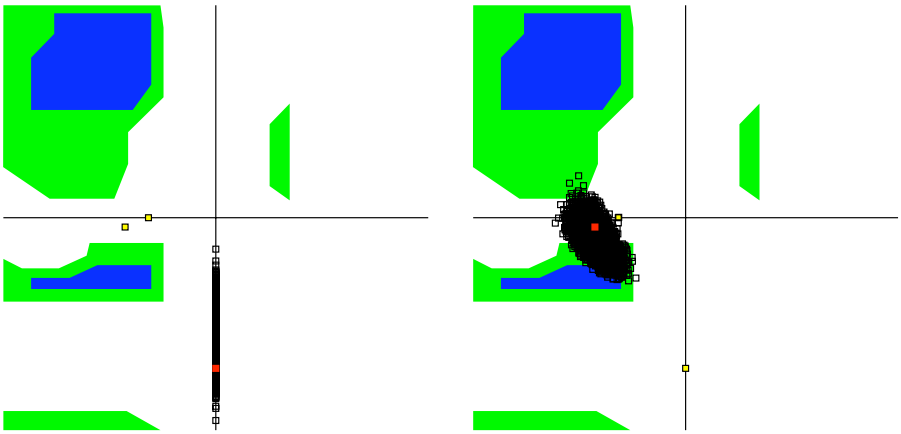
In the following analysis we consider the four different configurations of the GPE molecule discussed above, as a single molecule in solution with water and a single sodium counter-ion, approximating an infinite-dilution system, although periodic boundary conditions mean that this not technically accurate. These are referred to as A, B, C and D in the following discussion.

5.3.1 Simulation details

The boxes containing a single GPE and a single sodium ion in water were created using the configuration of the GPE molecule obtained from the end of the in *vacuo* simulation. This was placed, with a sodium ion, in a cubic box with sides of length 40 Å. This was repeated for each of the four conformational isomers, A, B, C and D, under consideration. A 40 Å box of water, containing 1500 water molecules of the SPC/E geometry [82], which had been previously subjected to a 2 ns MD simulation using the water potential was used to surround the GPE molecule and sodium ion with water. Water molecules that overlapped with GPE and sodium ions were removed. The four systems consist of a single GPE molecule, a sodium ion and 1412 water molecules.

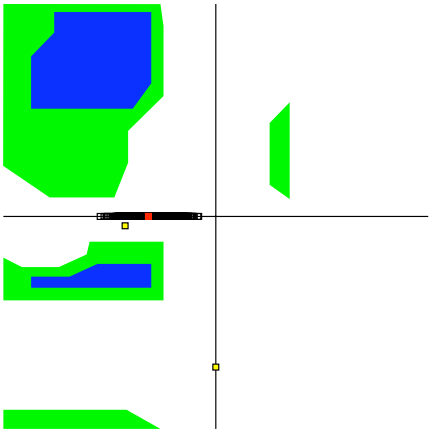
Each system was then subjected to a 200 ps NPT MD simulation at atmospheric pressure, 300 K and using the Berendsen thermo/barostat, using PINY_MD with the CHARMM22 and SPC/E force-fields. After the 200 ps NPT simulation a stable volume had been obtained for each system, resulting in each system being in a cubic box of length 35.0 Å.

The systems were then subjected to a 1 ns NVT equilibration simulation at 300 K, using the Nosé-Hoover thermostat, and a Nosé-Hoover chain of length 2. Data for the following discussion was then obtained from a further 1 ns production run. For the equilibration and production run simulations a time-step of 1 fs was used, with data collected every 100 time-steps (100 fs).

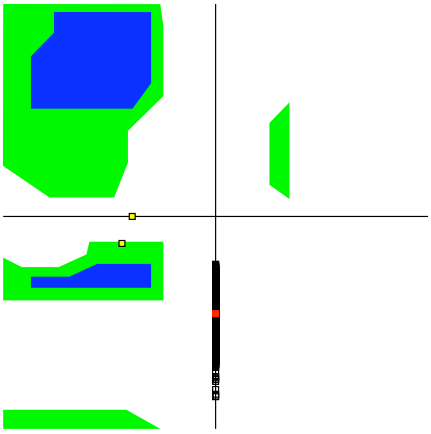


(a) Structure A: gly

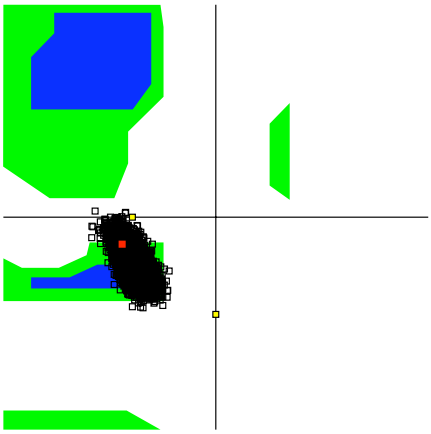
(b) Structure A: pro



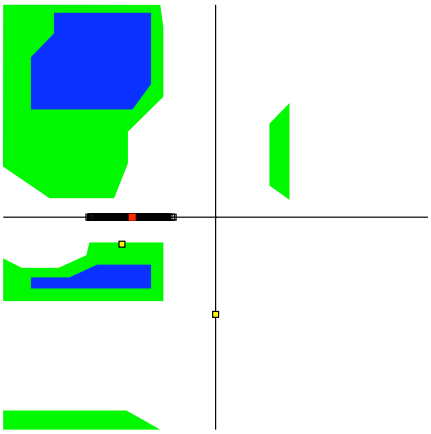
(c) Structure A: glu



(d) Structure B: gly

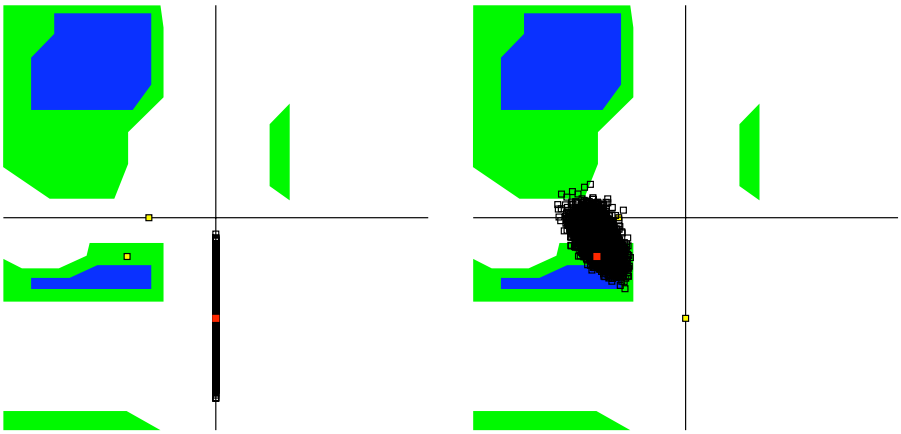


(e) Structure B: pro



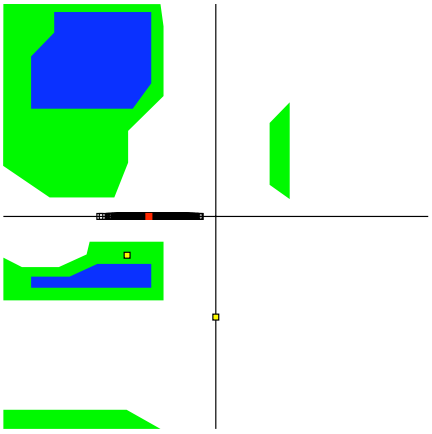
(f) Structure B: glu

Figure 5.8: Time-resolved Ramachandran maps for each torsional angle of the tripeptide structures A and B, simulated in a box of water with a single GPE molecule and a single sodium ion.

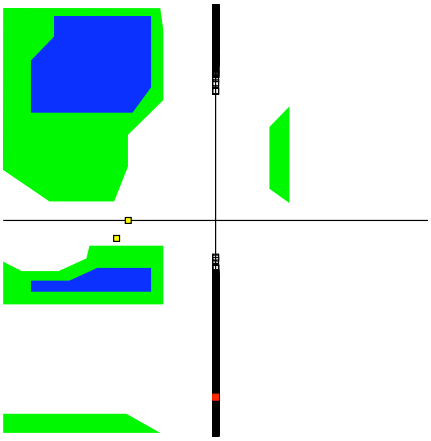


(a) Structure C: gly

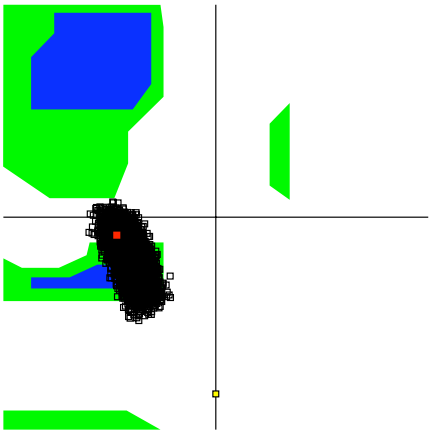
(b) Structure C: pro



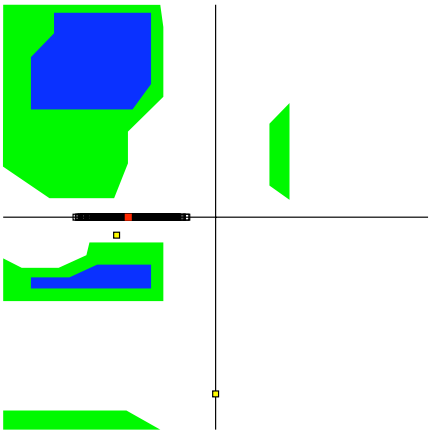
(c) Structure C: glu



(d) Structure D: gly



(e) Structure D: pro



(f) Structure D: glu

Figure 5.9: Time-resolved Ramachandran maps for each torsional angle of the tripeptide structures C and D, simulated in a box of water with a single GPE molecule and a single sodium ion.

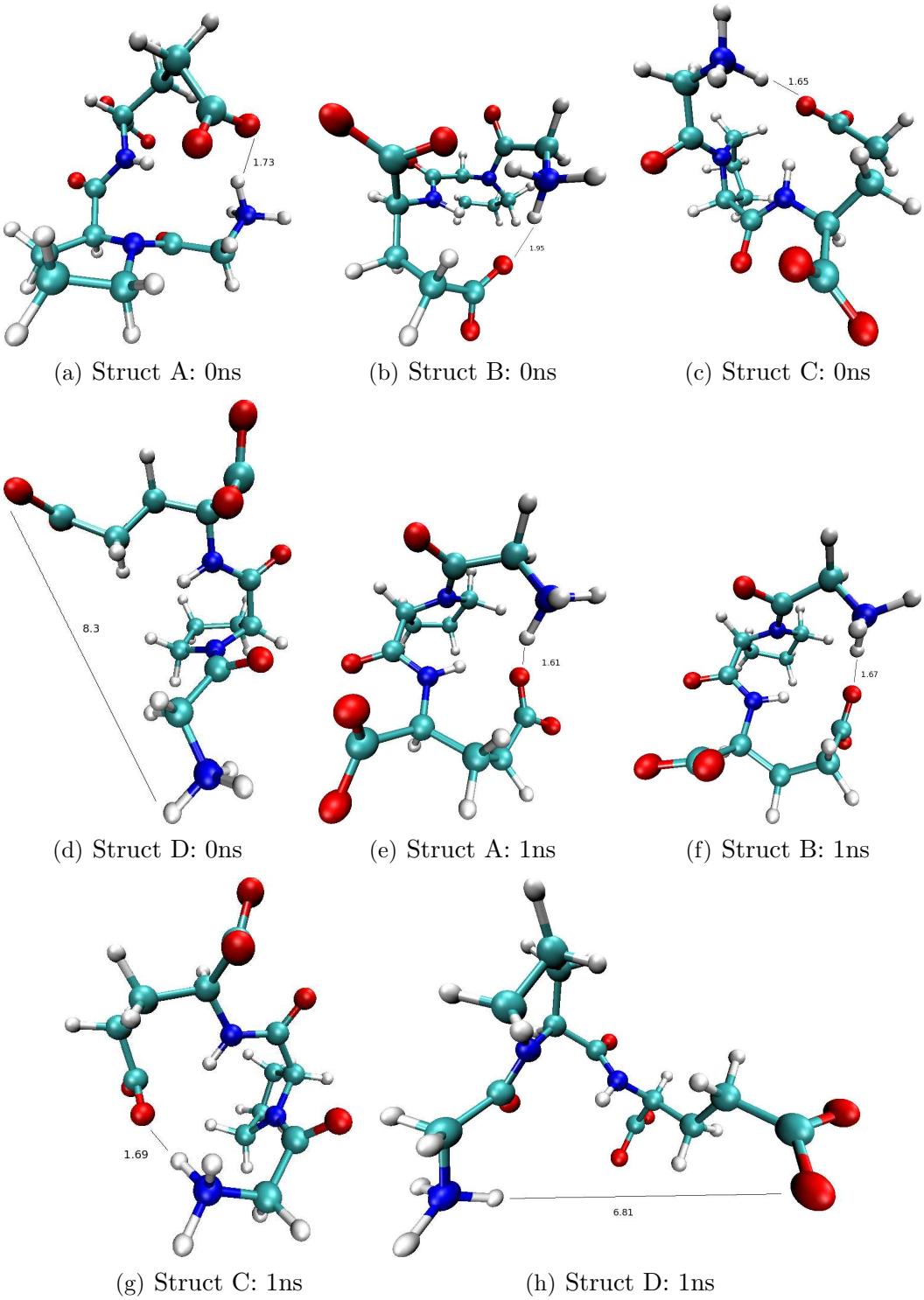
5.3.2 Orientation of the GPE molecule

We first consider the orientation of the tripeptide backbone using Ramachandran maps. The preliminary results discussed in section 5.2.1 on the orientation of the molecules suggested that the structures would quickly resolve into a similar orientation. Figures 5.8 and 5.9 show the time-resolved plots for the first 10 ps of the simulation trajectories (100 of the sampled time-steps), as this is representative of the entire trajectory. The three torsional angles are plotted, time-resolved, in three individual maps for each system along with the final position of the other two angles. In each case the red dot is the final position of the angle from the 10 ps sample.

The Ramachandran maps show broadly similar structures in the pro and glu torsional backbone angles. The only marked difference is in the gly 'twist', where structure D is the only molecule to still exhibit a positive ψ (y-axis) torsion angle. To understand this difference figure 5.10 shows snapshots of the conformation of the GPE molecule from each system taken at the beginning and end of the production run. Structures A, B and C all finish with 'folded' GPE molecules, where the amine group and the carboxyl group on the side-chain of the glu-molecule attempt to form an intra-molecular bond. Only structure D remains in an 'open' configuration.

In structure D the backbone of the GPE molecule remains approximately linear. Structure A is initially 'folded' and maintains an interaction between the same carboxyl and amine atoms throughout the entire 1 ns simulation. Structures B and C form tightly bound hydrogen bonds between the amine and carboxyl at the start and end of the simulation between the glu and gly chains, but unlike structure A, swap the pairs of atoms that are bonded. Structure B changes which O_{CE} atom (glu side-chain-oxygen) is involved in the bond, and structure C changes which H_X atom (gly amine-hydrogen) is involved in the bond. Given the swift change (in less than 10 ps) between bonding pairs, and our previous work on the glutamate molecule in chapter 3, these results suggest a bifurcated motif may exist to aid transition between the bonds.

Clearly we have two different types of structure, 'open' and 'closed' (also referred to as 'folded'). This reflects the difference in the Ramachandran maps where structure D, the only 'open' structure, had a significantly different gly-torsional angle compared to structures A, B and C. We further consider the



(a) Struct A: 0ns (b) Struct B: 0ns (c) Struct C: 0ns

(d) Struct D: 0ns (e) Struct A: 1ns (f) Struct B: 1ns

(g) Struct C: 1ns (h) Struct D: 1ns

Figure 5.10: Snapshots of the configuration of GPE structures A, B, C and D at the start and end of 1 ns production run.

structures from the end point of simulations A and D. We also consider the configuration of from the beginning of simulation A, where the GPE molecule was still 'open'. These three configurations of the GPE molecule, 'A open', 'A closed' and 'D open', are considered in aqueous solution simulations in the following section.

5.4 The structure of aqueous GPE solutions

The previous section, considering the simulation of a single GPE identified three configurations of the GPE molecule that warrant further study. This section considers the structure of aqueous GPE soluations using these three configurations, 'A open', 'A closed' and 'D open'.

5.4.1 Concentrations

The single GPE molecule simulations identified the key difference in structure being the amount that the molecule 'folded' up, forming an intra-molecular bond between the glu and gly residues of the molecule To assess whether this is an effect of solution concentration we have performed simulations on the molecules at two different concentrations; GPE:sodium:water at 1:1:29 and 1:1:61. In particular, considering the A 'open' and 'closed' configurations of the GPE molecule at two different concentrations should enable us to assess the role of water molecules in the folding of GPE molecules. We use the configuration of the GPE molecule obtained at the end of the single-molecule dilution simulations for A 'closed' and D 'open', and the configuration of structure A from the end of the in *vacuo* simulation for A 'open'.

The 1:1:29 (GPE:sodium:water) concentration is the same concentration used in the glutamate study. We use 34 GPE molecules, 34 sodium ions and 986 water molecules in a cubic box of length 34 Å. The second concentration of 1:1:61 (GPE:Sodium:Water) consists of 18 GPE molecules, 8 sodium ions and 1098 water molecules in a box of length 33 Å. Due to the size of the GPE molecule, the 1:1:61 concentration has a proportion of non-water atoms (rather than molecules) closer to the proportion of non-water atoms in 1:1:29 glutamate simulation.

In the following discussion we refer to the following systems:

- 1:1:61 closed A system: simulation at a concentration of 1:1:61, using the configuration for GPE from the end of the simulation A.
- 1:1:61 open A system: simulation at a concentration of 1:1:61, using the configuration for GPE from the end of the in *vacuo* simulation of structure A.
- 1:1:29 open D system: simulation at a concentration of 1:1:61, using the configuration for GPE from the end of simulation D.
- 1:1:29 closed A system: simulation at a concentration of 1:1:29, using the configuration for GPE from the end of simulation A.
- 1:1:29 open A system: simulation at a concentration of 1:1:29, using the configuration for GPE from the end of the in *vacuo* simulation of structure A.
- 1:1:29 open D system: simulation at a concentration of 1:1:29, using the configuration for GPE from the end of simulation of structure D.

5.4.2 Simulation Details

Each of the systems studied in this section were subjected to the same set-up, equilibration and production procedures as discussed in section 5.3.

The relevant GPE molecule and sodium ions were replicated on a lattice in a cubic box with sides of length 70 Å. In the 1:1:29 solution 34 GPE molecules and 34 sodium ions were placed on a lattice, for the 1:1:61 solution 18 GPE molecules and 18 sodium ions were placed on a lattice. A 70 Å box of water, constructed using the SPC/E geometric configuration and subjected to a 2 ns equilibration MD simulation using the SPC/E water potential, was used to surround each GPE system in water, removing water molecules that overlapped with GPE molecules and sodium ions, and such that the concentrations resulted in 1:1:29 and 1:1:61 (986 and 1098 water molecules remaining respectively).

Once the initial configuration was constructed, each system was subjected to a 200 ps NPT Berendsen thermo/barostat simulation, at atmospheric pressure and 300 K, using the PINY_MD simulation package. This treated the GPE molecules and sodium ions using the CHARMM22 forcefield parameters and water with the

SPC/E water-potential. After the 200 ps NPT simulation a stable volume had been obtained for each system. This resulted in each system having a cubic box of length of between 33 Å and 34 Å (the actual boxlengths at the end of the NPT simulation and used throughout the following NVT simulations are given in table 5.4).

System	Cell edge-length (Å)
1:1:61 A Closed	33.72
1:1:61 A Open	33.79
1:1:61 D Open	33.61
1:1:29 A Closed	34.32
1:1:29 A Open	34.36
1:1:29 D Open	34.28

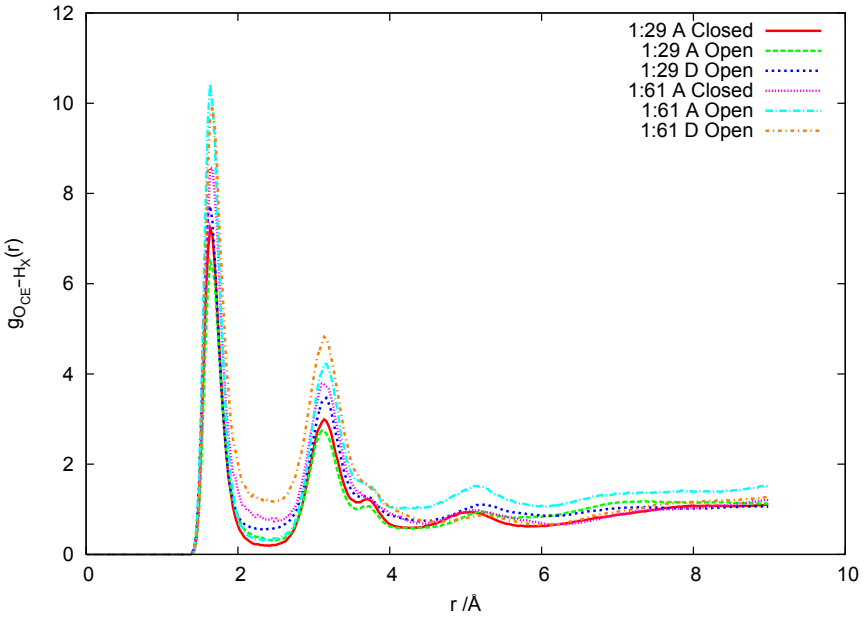
Table 5.4: The cubic box length of the post-NPT simulations of each of the GPE systems (used in the NVT equilibration and production runs).

An 2 ns equilibration simulation was performed, using the Nosé-Hoover NVT thermostat and Nosé-Hoover chains of length 2, at 300 K and atmospheric pressure and 1 fs time-steps. Data for the following discussion was then obtained from a further 2 ns production run, using the same variables. The data obtained for analysis was collected every 100 time-steps (100 fs), resulting in a trajectory of 20,000 configurations.

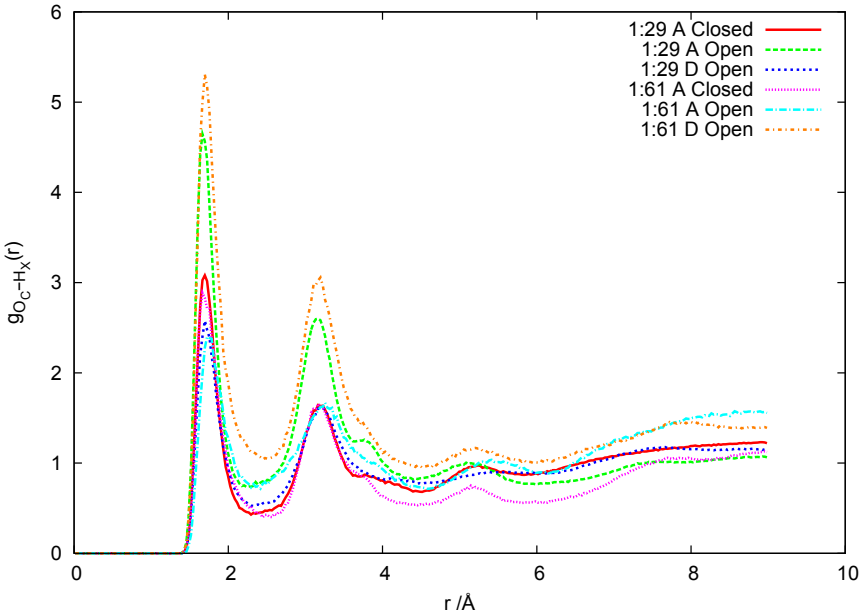
5.4.3 GPE-GPE interactions

Figures 5.11 and 5.12 show the RDFs between the carboxyl-oxygen atoms (O_{CE} and O_C) on the glu residue and the H_X and the N atoms of the amine on the glycine residue of GPE, and the corresponding glutamate interaction from chapter 3. The greater intensity first-peak of the $H_X - O_{CE}$ and $N - O_{CE}$ RDFs indicates a preference for GPE-GPE bonds to form *via* the side-chain of the glu-residue rather than the backbone carboxyl (OC). This is verified by the co-ordination num bers shown in table 5.5, which show that all six systems have a greater $H_X - O_{CE}$ co-ordination than $H_X - O_C$ co-ordination number. This is the same as the preference observed for bonds to form between the side-chain and the amine in the glutamate solutions as discussed in chapter 3.

In both the $g_{H_X-O_{CE}}(r)$ and the $g_{H_X-O_C}(r)$ plots, figure 5.11, the peaks lie

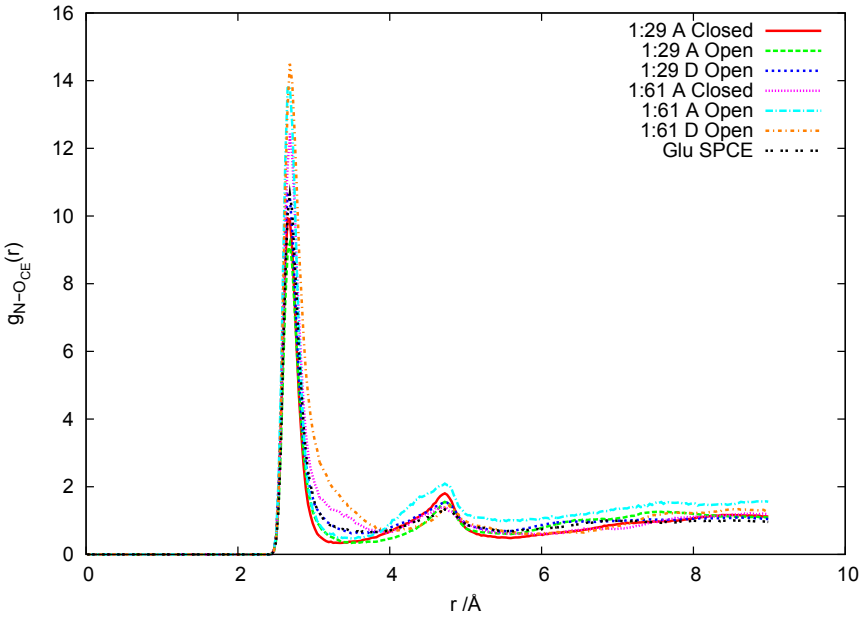


(a) $g_{H_X-O_{CE}}(r)$

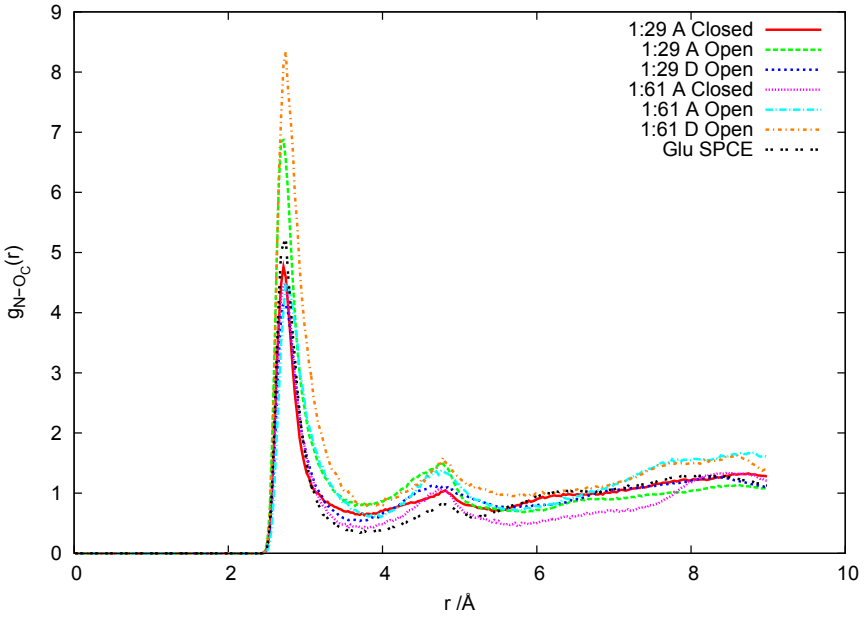


(b) $g_{H_X-O_C}(r)$

Figure 5.11: Radial distribution functions for the GPE-GPE interactions in 1:1:29 and 1:1:61 solutions of GPE.



(a) $g_{N-O_{CE}}(r)$



(b) $g_{N-O_{CE}}(r)$

Figure 5.12: Radial distribution functions for the GPE-GPE interactions in 1:1:29 and 1:1:61 solutions of GPE.

	1:1:29			1:1:61		
	A closed	A open	D open	A closed	A open	D open
$n_{O_{CE}}^{H_X}$	0.251	0.242	0.279	0.190	0.193	0.237
$n_{O_C}^{H_X}$	0.147	0.211	0.137	0.084	0.0813	0.156
$n_{O}^{H_X}$	0.0547	0.114	0.0329	0.302	0.0467	0.00739
$n_{O_{CE}}^{N}$	0.789	0.784	0.946	0.634	0.612	0.839
$n_{O_C}^{N}$	0.524	0.800	0.528	0.302	0.342	0.598
n_{O}^{N}	0.253	0.494	0.170	0.0595	0.197	0.0660

Table 5.5: Co-ordination numbers, n_β^α, for the GPE-GPE hydrogen bond sites, indicating the average number of β atoms around an α atom. For comparison purposes we have used a cut-off distance of 2.2Å for the $O_{CE}/O_C/O - H_X$ co-ordination numbers and a cut-off distance of 3.2Å for the $O_{CE}/O_C/O - N$ co-ordination numbers.

in broadly similar positions for each of the systems shown. As identified in the parallel-tempering studies of glutamate, there is a shoulder in the trailing edge of the second peak for the GPE systems. This shoulder is most prevalent in the high concentration solutions (1:1:29) and between the O_C and H_X atoms. As with the classical, non-parallel tempering glutamate RDFs, there is only the 'suggestion' of a shoulder in the lower concentration solutions of the $O_C - H_X$ RDFs and at all concentrations in the $O_{CE} - H_X$ RDFs. This indicates that the motif causing this shoulder exists between both species of carboxyl group and the amine group, but is more common at high concentrations and between the C_α-carboxyl and the amine, than the side-chain carboxyl and the amine. In the glutamate analysis we attributed this peak to the presence of the three hydrogen atoms in the amine group all interacting with the carboxyl-oxygens (see diagram in figure 4.15). It should be noted that the shoulder's position, between 3.6 Å and 4.2 Å, is consistent with a water molecule mediating the interaction between two GPE molecules. Observation of the trajectories of the molecules over time enabled visual confirmation of both motifs. A snapshot of the GPE-water-GPE interaction is shown in figure 5.13.

To further assess interactions between the carboxyl oxygens and the amine groups of different GPE molecules, we consider 2D radial-radial distribution functions for the carboxyl-amine interactions. In figure 5.14 we show the plots correlating the distances between one H_X atom and each of the two oxygen

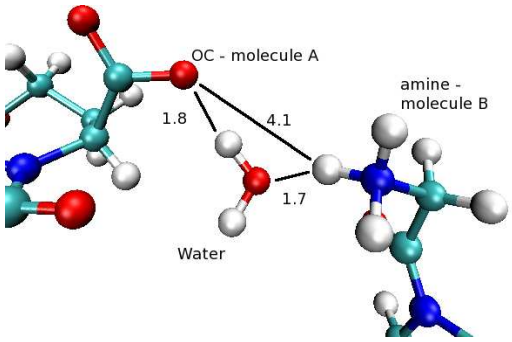
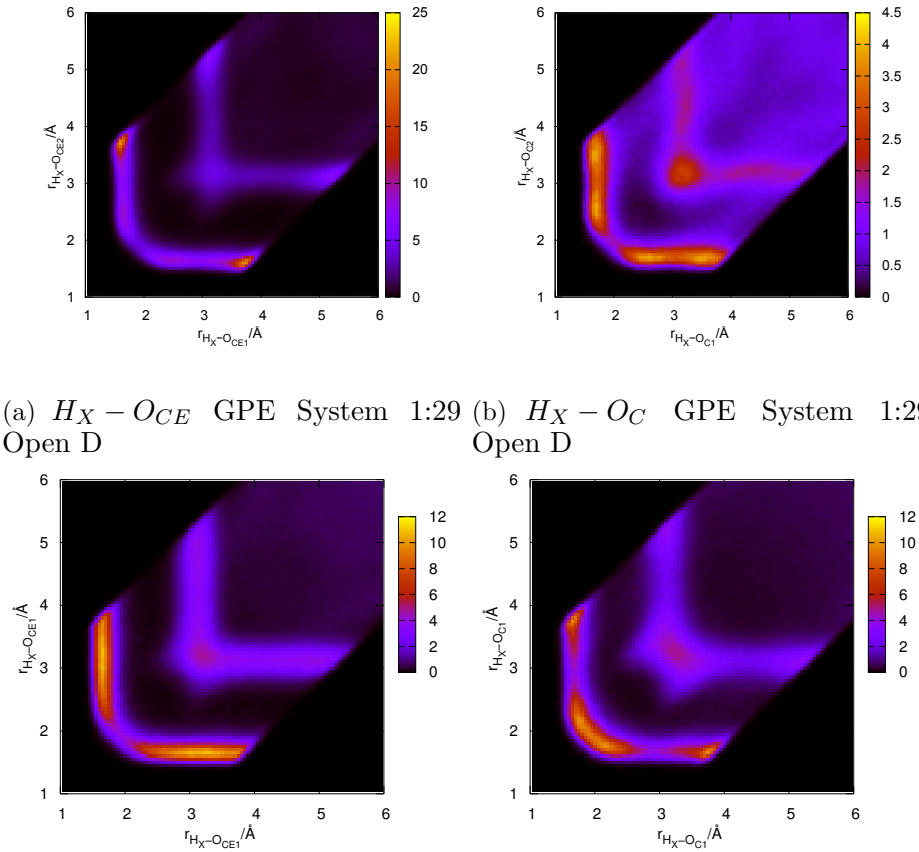


Figure 5.13: Snapshot of system 1:1:61 Open A, showing the interaction between a carboxyl group on molecule A with the amine group on molecule B, mediated by the presence of a water molecule.

atoms from the same carboxyl group for both the GPE-GPE and the glu-glu interactions. In both GPE plots the non-zero intensity at $r_{H_X-O_{C1}} \simeq 1.8$ Å, or $r_{H_X-O_{C2}} \simeq 1.8$ Å, indicates single, linear bonds, and the non-zero intensity at $r_{H_X-O_{C1}} \simeq r_{H_X-O_{C2}} \simeq 1.8$ Å, is consistent with a bifurcated motif.

Comparison with the glu-glu interactions, from the classical glutamate study in chapter 3, show a marked difference in the intensity of the plots for glu and GPE. The two types of amine-carboxyl interaction in the glu simulations produced broadly similar intensities around the single and bifurcated bonding regions in each of the glutamate-glutamate 2D radial-radial plots. In contrast, in the GPE simulation the scales for the plots of the two types of amine-carboxyl bonds are very different, with the $H_X - O_{CE}$ intensity peaking at five times that of the $H_X - O_C$ intensity. Additionally, the greatest intensity in the $H_X - O_{CE}$ plot occurs around the region produced by single bonds, where $r_{H_X-O_{CE1}} \simeq 1.8$ Å or $r_{H_X-O_{CE2}} \simeq 1.8$ Å, and not for bifurcated bonding motifs. In the $H_X - O_C$ plot for the GPE-GPE interaction (figure 5.14(b)), there is no clear preference for linear or bifurcated bonds, while the corresponding glu-glu interaction (figure 5.14(d)) shows a clear preference for bifurcated bonds.

In summary, we find that adjacent GPE molecules in general prefer to form inter-molecular bonds between the amine of the glycine residue with the side-chain carboxyl of the glutamate residue rather than the C_α-carboxyl. The number of bifurcated bonds formed between the amine group and each of the carboxyl groups on GPE molecules is similar, although as a percentage of the total bonds formed on each of the carboxyls, the C_α-carboxyl has no clear preference to form single



(a) $H_X - O_{CE}$ GPE System 1:29 Open D

(b) $H_X - O_C$ GPE System 1:29 Open D

(c) $H_X - O_{CE}$ Glu

(d) $H_X - O_C$ Glu

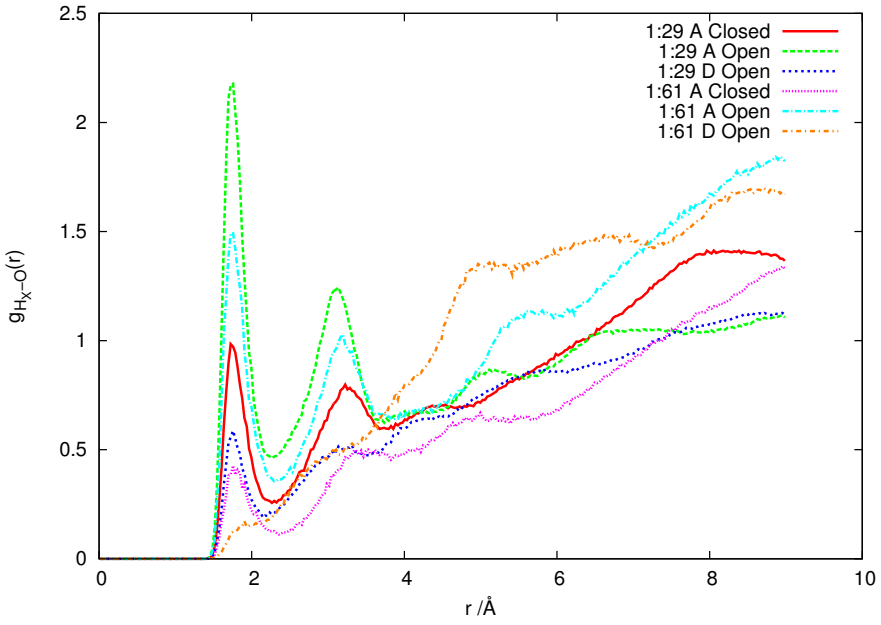
Figure 5.14: 2D radial-radial distribution functions for the GPE-GPE interactions between H_X and O_{CE}/O_C (for the 1:1:29 D open system) and the corresponding glutamate plots.

or bifurcated motifs, but the side-chain clearly prefers to form single bonds.

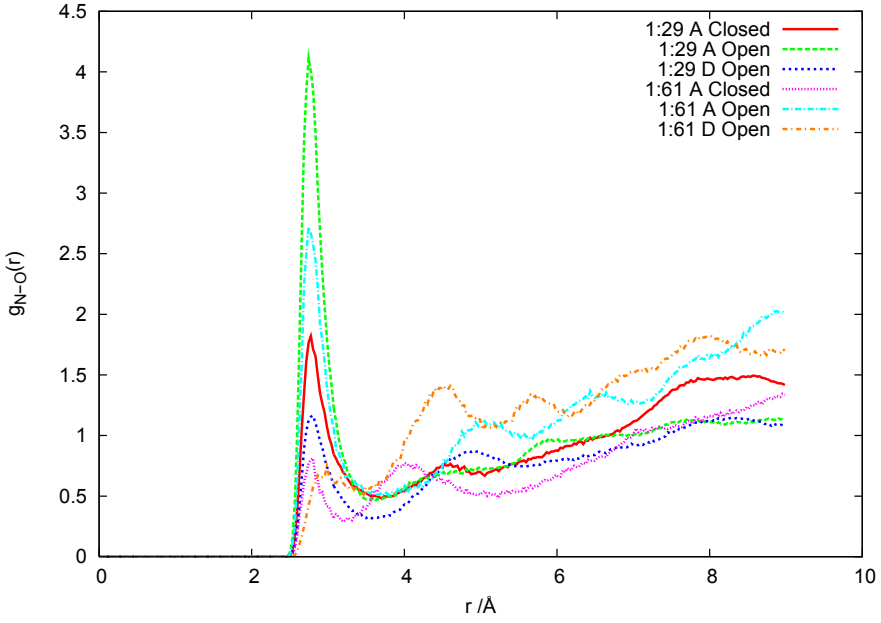
The other five GPE systems all produced similar 2D radial radial distribution plots for the amine-carboxyl interactions. These are shown in Appendix C, figures C.1 and C.2. We also checked whether an oxygen atoms was shared between and H_X atom, forming a different type of bifurcated motif. As observed in the glutamate study, the 2D radial-radial distribution functions correlating O_{CE}/O_C with two H_X atoms (plotting $r_{O_{CE}/O_C-H_{X1}}$ against $r_{O_{CE}/O_C-H_{X2}}$, figures C.3 and C.4, showed zero intensity in the bifurcated region, indicating no bifurcated motifs of this type.

GPE molecules, unlike glutamate molecules, can also form inter-molecular bonds*via*the peptide backbone. The pro and gly residues both have a negatively charged oxygen atom as part of the peptide bonds. These atoms, labelled 'O' in the labelling scheme shown in figure 5.1, can form inter-molecular bonds with the amine groups of adjacent GPE molecules. In figure 5.15 we show the radial distribution functions correlating the$O$ atoms of the peptide bonds to the H_X and the N atoms of the gly-amine. Here we see a well-defined first peak followed by a deep trough, indicating a hydrogen bond forming between the amine groups and the peptide backbone oxygens. The second peak, at $\simeq$3 Å, in the $H_X - O$ RDF is consistent with the correlation of the other two H_X atoms in the amine group (not directly involved in the bond) with the O atom.

The first peak intensities of the $H_X/N - O$ RDFs is a fifth and a tenth of the intensity of the first peak of the O_C and O_{CE} RDFs respectively, indicating a strong preference for carboxyl-amine GPE-GPE interactions. Thecoordination numbers suggest a preference for GPE-GPE bonds forming *via* the O_{CE} atoms, with all but the 1:1:29 A Open system having double the coordination around the side-chain compared to the C_α-carboxyl. In addition we see that the RDF continues to rise in intensity above 6Å, suggesting that there is a disproportionately small number of$O$-amine interactions in the small-r region, resulting in the RDF intensity being unusally small until reaching the normalisation value of ~ 1 at large r. This effect is greatest in the lower 1:1:61 concentration systems (the torquoise, yellow and pink lines of figures 5.15(a) and 5.15(a)). As GPE-GPE bonds*via*the peptide backbone are not favoured, this sharp rise in the 1:1:61 systems is consistent with fewer GPE molecules in close proximity*via*the amine-backbone interaction, further indicating that this is not a



(a) $g_{H_X-O}(r)$



(b) $g_{N-O}(r)$

Figure 5.15: Radial distribution functions for the GPE-GPE interactions in 1:1:29 and 1:1:61 solutions of GPE.

favoured configuration.

Figure 5.16 shows the 2D radial-radial distribution functions correlating the distances between the O atoms of the peptide backbone with two of the amine H_X atoms in the 1:1:29 open D system. Similar motifs were found in the other five systems, as shown in figure C.5.

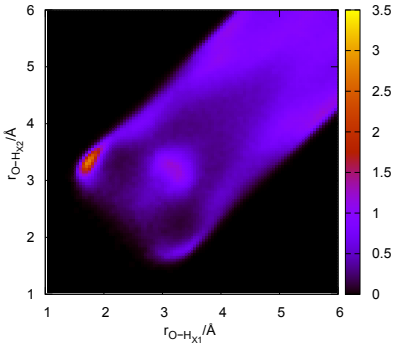


Figure 5.16: 2D radial-radial distribution functions for the GPE-GPE interactions $O - H_X$ (for the 1:1:29 D open system). The plots for the other GPE systems are shown in Appendix C, figure C.5.

The zero intensity in the region $r_{O-H_{X1}} \simeq r_{O-H_{X2}} \simeq 1.8$ Å indicates that no bifurcated bonds exist between O and H_X atoms, where two H_X atoms could simultaneously bond to the same O atom. This is consistent with the lack of carboxyl-amine bifurcated bonds of this type. The only carboxyl-amine bifurcated motifs that exist involve two carboxyl oxygens being shared between one H_X atom. As this type of bifurcated bond can not occur with the backbone oxygens, the inability of the backbone to form bifurcated bonds may explain the smaller proportion of GPE-GPE bonds*via*the O-atoms. We suggested in the glutamate results that the bifurcated bond could provide a lower-energy pathway between single linear bonds. Removing this route between bonds means that there is a greater energetic impediment to forming new single linear bonds.

Table 5.5 shows the co-ordination numbers for the GPE-GPE bonding interactions. This shows the number of oxygen atoms occupying the bonding region between around the H_X and N atoms, averaged across all atoms of that particular species over the entire simulation. A radial cut-off for 'bonded' H_X-oxygen atoms was taken as 2.2 Å. This was the average of the H_X-oxygen (for

GPE-GPE interactions) RDF first-minima across all three types of oxygen and all six systems modelled. Similarly the radial cut-off for bonded N-oxygen atoms was taken as 3.2 Å. The average position of the individual first minima, the corresponding $g(r)_{\alpha\beta}$ and the occupation number at the individual first minima, are shown in table C.1 in appendix C.

The greatest GPE-GPE occupation number, in all six systems, occurs for the amine-side chain interaction: $n^{H_X}_{O_{CE}}/n^{N}_{O_{CE}}$, and the smallest for the amine-backbone oxygen interaction: $n^{H_X}_{O}/n^{N}_{O}$. These results are consistent with the difference in the first peak intensities of the corresponding RDFs in figures 5.11,5.15. This is also consistent with our findings for the aqueous glutamate solutions, where glu-glu bonds were more likely to occur *via* the side-chain carboxyl. This was attributed in part to the flexibility of the side-chain compared to the C_α carboxyl, which is much more tightly constrained. This applies to the GPE molecule as well, as the peptide backbone is far more tightly constrained than the side-chain of the glu-residue.

Other explanations for the preference to bond *via* the side-chain of the glu residue are also similar to those discussed for the glutamate simulations. Firstly, the O_{CE} atoms have a greater negative potential (O_{CE} has charge of -0.76 eV compared to -0.67 eV for the O_C atoms). Secondly, in the glutamate simulations the C_α-carboxyl is in closer proximity to the amine group within the same molecule, reducing its ability to attract other amine groups, as approaching molecules are more susceptible to any electronegative repulsion the amine groups may exert on each other, than if the bond formed *via* the side-chain. In the GPE molecules this effect is likely to be negligible as with three residues the proximity between the C_α-carboxyl and the amine within the same molecule is not significant. Thirdly, the side-chain of the glutamate residue, as discussed for the glutamate systems, has greater flexibility than the backbone. The backbone torsional angles tightly constrain the position of the C_α-carboxyl, limiting the orientational freedom of the group, and therefore the ability of the O_C atoms to form bonds with adjacent molecules compared to the O_{CE} atoms on the side-chain.

The occupation numbers are higher for the amine-carboxyl interactions in the 1:1:29 systems compared to the 1:1:61 systems, except for the Open D systems. The higher concentration of GPE in the 1:1:29 systems may result in

more GPE-GPE bonds, and therefore greater occupation numbers. This is only reflected by the amine-carboxyl occupation numbers. The amine-O interaction has a lower occupation number at high concentration. Higher concentrations of GPE molecules may inhibit the movement of GPE molecules as there is less freedom of movement against other large molecules, and therefore there is a greater energetic impediment to forming and breaking inter-molecular bonds. The amine-O occupation number is smaller than the amine-carboxyl occupation number in the high concentration systems, also suggesting that this type of interaction is more susceptible to configurational constraints. As mentioned above, this may also be due to the inability of the backbone oxygens to form bifurcated bonds. Bifurcated bonds between water molecules have been shown to have approximately half the energy of a single bond [65]. The inability to form bifurcated bonding motifs may inhibit the ability to form inter-molecular bonds along the GPE backbone, particularly with increasing concentration where the energetic barrier to forming bonds is increased due to the lack of orientational freedom.

The occupation number for $n_{O_{CE}}^{H_X}$ and $n_{O_C}^{H_X}$ is similar for both concentrations of the D Open systems, indicating that the prefered type of GPE-GPE bonds is not dependent on concentration in this case. The D Open system uses a configuration for the GPE molecule that did not form a closed structure for a significant proportion of the initial single-molecule simulation. One possible explanation for the difference in bonding prevalance between the D Open systems and the A Open/A Closed systems is the number of intra-molecular bonds, resulting in 'closed' motifs, which presumably reduce the ability to form inter-molecular bonds on these sites.

Table 5.6 gives the number of GPE molecules in the final step of each simulation that are 'closed', forming an intra-molecular bond between one of the carboxyl groups and the amine group. (No intra-molecular O-amine bonds were found.) As with the inter-molecular occupation numbers, we used a radial-cutoff distance of 2.2 Å to define the $H_X - O_{CE}/O_C$ bond.

All six systems have formed intra-molecular bonds in GPE molecules at the very end of the production run, despite the Open A and Open D systems having an initial configuration where all the GPE molecules were fully extended, with no intra-molecular bonding. More 'closed' than 'open' molecules were found in the

System	GPE molecules	Closed Molecules $O_C - H_X$	$O_{CE} - H_X$	Total	Open Molecules
1:1:29 Closed A	34	21	2	23	11
1:1:29 Open A	34	19	2	21	13
1:1:29 Open D	34	7	6	13	21
1:1:61 Closed A	18	11	0	11	7
1:1:61 Open A	18	2	11	13	5
1:1:61 Open D	18	5	1	6	12

Table 5.6: The number of GPE molecules in a closed configuration, forming intra-molecular $O_C - H_X$ or $O_{CE} - H_X$ bonds, and the number of GPE atoms that remain unbonded at the end of the production simulation.

Closed and Open A systems at both concentrations. This trend is reversed for the Open D system, where more open molecules were found at both concentrations. As all six systems, irrespective of their starting configuration have some closed and some open molecules, we can conclude that both configurations are energetically favoured in certain circumstances. The 'A' systems prefer closed motifs, even in the 'open' A initial configuration system. This supports the findings of the single-molecule dilution case where the 'open' A molecule quickly closed up. Similarly the greater proportion of open molecules in the two 'D' systems follows the observation of the an open molecule throughout the initial single-molecule in solution simulation.

There is a clear preference to form intra-molecular bonds*via* the C_α-carboxyl than*via* the side-chain across all six systems. This explains the propensity to form *inter*-molecular bonds*via* the side-chain in all six systems as the C_α carboxyl forms more intra-molecular bonds.

As this data only represents the configurations from the final step of the simulations it is not prudent to draw too many conclusions from these results. We have also calculated the proportion of 'open' and 'closed' GPE molecules throughout the entire trajectory, as shown in table 5.7. These results take the number of open and closed molecules at every time-step in the simulation, before averaging across all time-steps. This shows that the greatest proportion of molecules remaining 'open' is in the 1:1:29 Open D system, where 66% of the GPE molecular configurations, averaged across the entire simulation, remain free from intra-molecular bonds, reflecting the final-configuration results above. The

table also shows a similar number of 'open' molecules in both concentrations of the A closed systems, suggesting that for this initial A closed configuration, the concentration of GPE molecules does not affect the formation of intra-molecular bonds significantly.

System	% closed Molecules			% Open Molecules
	$O_C - H_X$	$O_{CE} - H_X$	Total closed	
1:1:29 Closed A	59.4%	9.89%	69.2%	32.8%
1:1:29 Open A	50.2%	8.55%	58.8%	41.2%
1:1:29 Open D	21.1%	13.2%	34.3%	65.7%
1:1:61 Closed A	65.7%	1.64%	67.3%	32.7%
1:1:61 Open A	3.82%	73.0%	76.8%	23.2%
1:1:61 Open D	12.1%	36.6%	48.7%	51.3%

Table 5.7: The proportion of GPE molecules in a closed loop, forming intra-molecular $O_C - H_X$ or $O_{CE} - H_X$ bonds, and the proportion of GPE atoms that remain unbonded averaged over the entire production simulation.

In all three 1:1:29 concentration systems GPE intra-molecular bonds prefer to form*via*the C_α-carboxyl instead of the side-chain carboxyl. At the lower concentration (the 1:1:61 systems), both the Open A and Open D systems switch preference to the side-chain for intra-molecular bonds. This should result in a lack of GPE-GPE bonds forming*via*the side-chain in the 1:1:61 Open A and Open D systems compared to the 1:1:29 systems. This is not reflected in the $n^{H_X}_{O_{CE}}$ occupation numbers. For the Open D system the occupation number, $n^{H_X}_{O_{CE}}$, is 0.237 and 0.279 for the 1:1:61 and 1:1:29 concentrations respectively; a small increase in the side-chain inter-molecular bonds forming with increasing concentration. A similar trend is seen in the Closed A system, with an increase from 0.190 to 0.251 with increasing concentration. This is the opposite trend to that expected if the increase in C_α-carboxyl intra-molecular bonds was to effect the inter-molecular bonds For the Open A system we see an increase in $n^{H_X}_{O_{CE}}$ occupation numbers from 0.193 to 0.242 with a decrease in concentration from the 1:1:61 to the 1:1:29 systems, indicating a small reduction in side-chain inter-molecular bonding. The increase in intra-molecular bonding*via*the side-chain, as concentration changes, also coincides with a small increase in inter-molecular co-ordination number.

If GPE molecules are forming intra-molecular bonds this reduces their ability

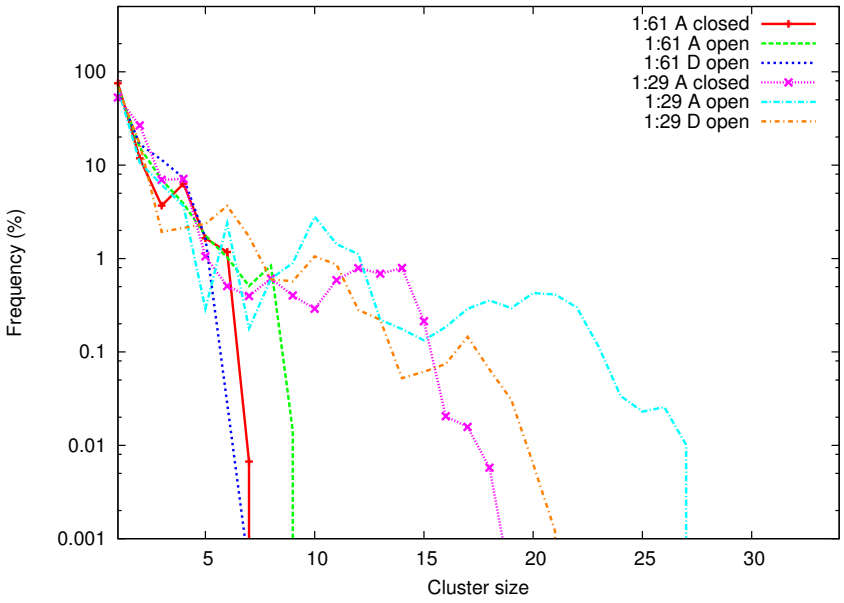


Figure 5.17: Distribution of GPE molecule cluster sizes as a fraction of the total clusters.

to form inter-molecular bonds, and therefore we expect that smaller clusters of GPE molecules would form where the proportion of 'closed' molecules is greatest. Figure 5.17 shows the distribution of cluster sizes in each of the system. This is taken as the proportion of clusters of size n, as a fraction of the total number of clusters found, averaged across all of the 20,000 trajectory snapshots from 2 ns production runs. The theoretical maximum possible cluster size would be 18 for the 1:1:61 systems and 34 for the 1:1:29 systems.

The distribution of cluster sizes shows that in all of the systems the GPE molecules prefer to remain as a single molecule. In all 6 systems there are a large proportion of the clusters involving up to 4 or 5 GPE molecules, but few clusters larger than this. As with the glutamate distribution we see no percolating clusters. We also see that the 1:1:29 systems that have 34 molecules have a greater proportion of large clusters than the 1:1:61 systems with just 18 molecules, indicating that large clusters are related to the overall number of GPE molecules and the concentration. At the higher concentration there is a relationship between the proportion of closed molecules and the largest cluster size obtained. The 1:1:29 A closed system had contained a largest cluster size of 17 members and had 69.2% of its molecules in a closed state, while the This trend is not continued

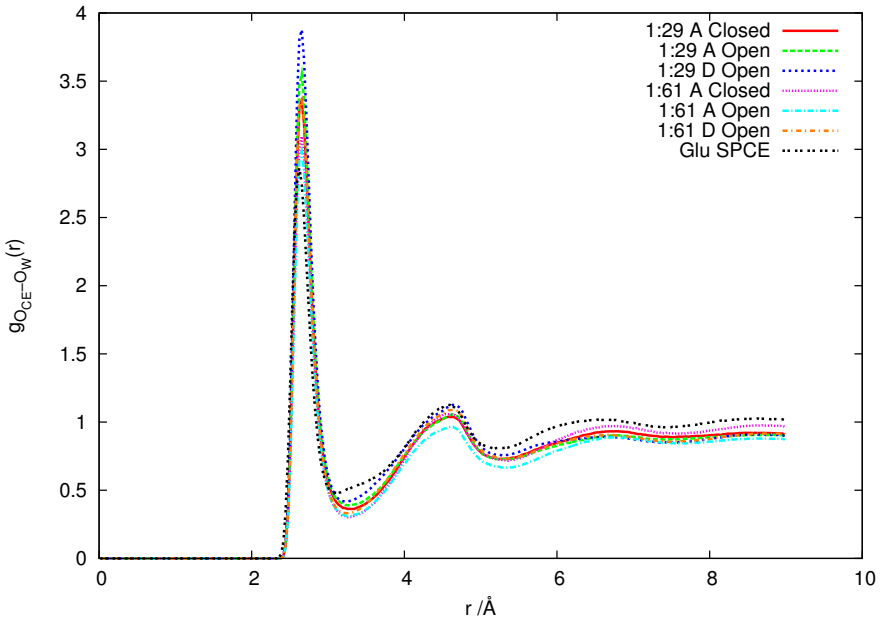
in the lower concentration systems, and the largest cluster obtained does not vary with the proportion of closed molecules even at the 1:1:29 concentration as the open D system with 34.4% of its GPE molecules in closed configuration does not contain a maximum cluster size larger than that of the open A system which has 58.8% of its molecules closed.

5.4.4 Water-GPE interactions

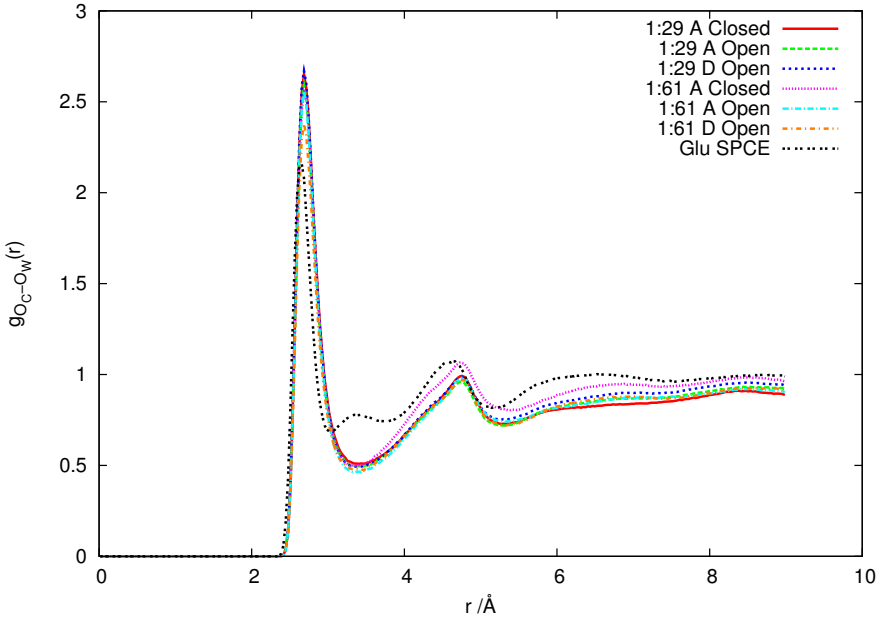
Figures 5.18 and 5.19 show the RDFs for the carboxyl oxygens of GPE interacting with water molecules. The clearly defined first peak in all four plots indicates that a hydrogen bond forms between both types of carboxyl group on the GPE molecules and water molecules, with an $H_W - O_C/O_{CE}$ bond cut-off distance of $\sim$2.2 Å, and an $O_W - O_C/O_{CE}$ bond cut-off distance of $\sim$ 3.3Å.

The second peak in the O_W-carboxyl plots in figure 5.18 is due to the second carboxyl oxygen atom co-ordinating with the water molecule. Similarly, the second peak in the H_W-carboxyl plots in figure 5.19 is due to the second hydrogen in the water molecule interacting with the carboxyl oxygen, and the second (non-bonded) carboxyl oxygen interacting with the bonded hydrogen. The shoulder on the trailing edge of the second peak in the H_W-carboxyl plots, between $r \simeq 3.5$ and $r \simeq 4.0$ Å, is due to the non-bonded carboxyl oxygen interacting with the non-bonded hydrogen.

There are no clear structural features differentiating the results of the six GPE systems in the water-GPE RDFs. The co-ordination numbers shown in table 5.8 show that in the 1:1:29 system there are more $O_{CE} - O_W$ bonds than $O_C - O_W$ bonds. In the 1:1:61 system the difference is minimal, and varies with the Open D system forming more side-chain bonds with water and the Closed A and Open D systems preferring to form GPE-water bonds *via* the C_α carboxyl. There is no clear difference between the two concentrations, indicating that although there was a slightly greater number of GPE-GPE bonds forming in the 1:1:29 concentration systems (see table 5.5) that any side-effects changing the behaviour of water is not significant. Biologically this may mean that mild concentration changes of peptides will not effect their bonding abilities, although the prevalence of GPE-GPE bonds at both concentrations does not answer whether the system slows down at higher concentrations, limiting the ability of GPE molecules to diffuse through a system. Further dynamical studies on GPE should be carried

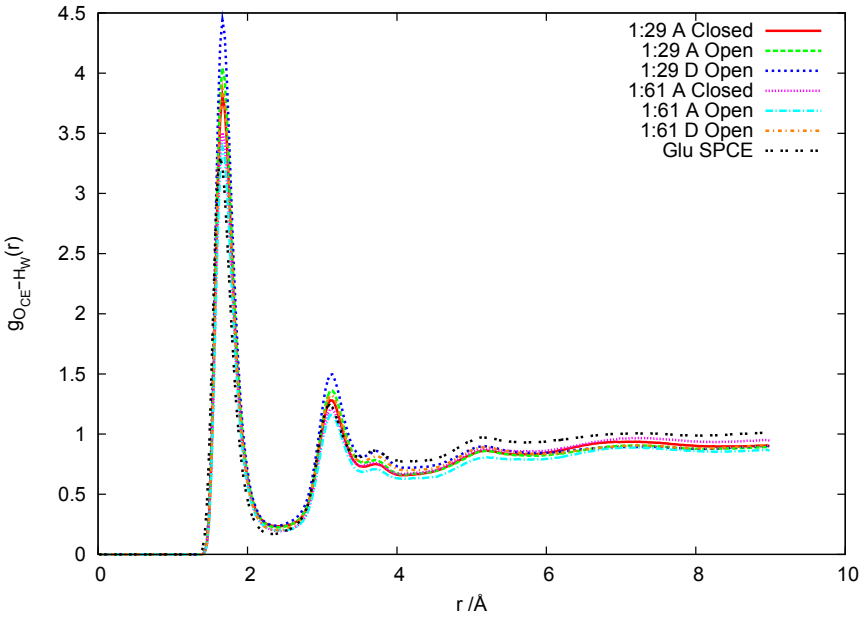


(a) $g_{O_W-O_{CE}}(r)$

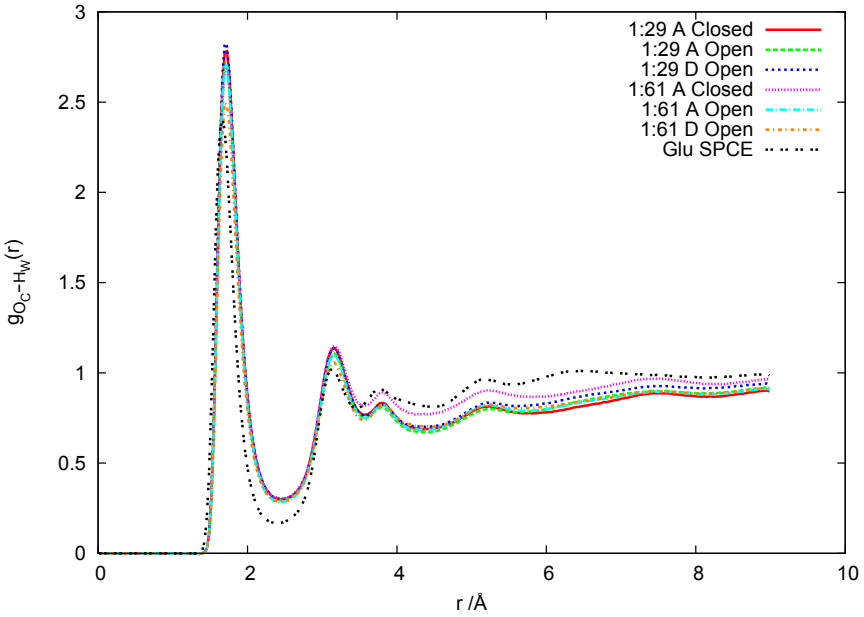


(b) $g_{O_W-O_C}(r)$

Figure 5.18: Radial distribution functions for the GPE-carboxyl and O_W-water interactions in 1:1:29 and 1:1:61 solutions of GPE.



(a) $g_{H_W-O_{CE}}(r)$



(b) $g_{H_W-O_C}(r)$

Figure 5.19: Radial distribution functions for the GPE-carboxyl and the H_W-water interactions in 1:1:29 and 1:1:61 solutions of GPE.

out to assess the bond lifetimes of GPE-GPE and GPE-water interactions.

System	$n_{O_W}^{O_{CE}}$	$n_{O_W}^{O_C}$	$n_{O_W}^{N}$
1:1:29 Closed A	2.73	2.69	1.43
1:1:29 Open A	3.12	2.69	1.50
1:1:29 Open D	3.12	2.69	1.62
1:1:61 Closed A	2.83	2.96	1.54
1:1:61 Open A	2.74	2.88	1.39
1:1:61 Open D	3.08	2.80	1.61

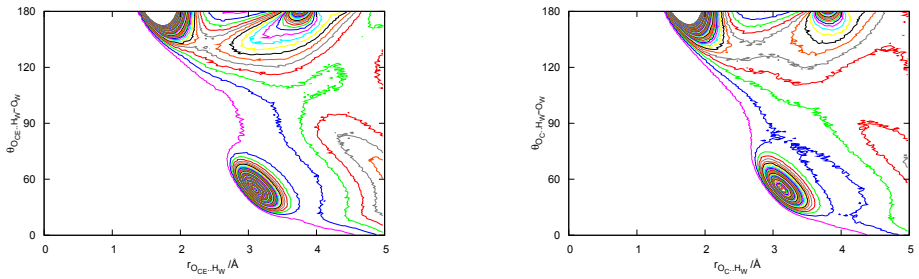
Table 5.8: Co-ordination numbers for the GPE-GPE interactions $O_{CE} - O_W$, $O_C - O_W$, and $N - O_W$ interactions.

All four plots also show the SPC/E water-glutamate RDFs, obtained from the simulations discussed in chapter 3. These show a more clearly defined first trough in the $H_W - O_C/O_{CE}$ RDFs at ∼2.2 Å compared to the GPE solutions, suggesting that the GPE system forms slightly weaker bonds with water than the glutamate molecules did.

The main difference between the glutamate-water bonds and the GPE-water bonds however is the difference in $g_{O_W-O_C}(r)$ in figure 5.18(b). The glutamate solution has a shoulder between the first and second peaks, attributed to a 'circle'-motif where the water molecules bridge the gap between the amine and C_α-carboxyl groups. This feature is absent from the GPE data, as the distance between the amine and the carboxyl on the peptide backbone is too large for a single water molecule to bridge. Here we see that the glutamate solution produced a shoulder on the leading-edge of the second peak at around 3Å (black dotted line).

Figure 5.20 shows the plots of the 2D radial-angular correlation between r_{O_{CE}/O_C-H_W} distance with the $\theta_{O_{CE}/O_C-H_W-O_W}$ angle for the 1:1:29 Open D system (similar plots were obtained for the other five systems, which are shown in appendix C, figures C.8 and C.9). Both plots show three clearly distinct peaks. The first at $r_{O_{CE}/O_C-H_W} \simeq 1.8$Å, $\theta_{O_{CE}/O_C-H_W-O_W} \simeq 180°$ corresponds to the linear hydrogen bond between the carboxyl group and water, involving one of the hydrogen atoms (H_{W1}). This indicates that the hydrogen bonds are constrained to a linear or nearly linear bond, fluctuating by less than 30° before the bond breaks and the molecules move apart. The second peak at $r_{O_{CE}/O_C-H_W} \simeq 3.3$Å, $\theta_{O_{CE}/O_C-H_W-O_W} \simeq 60°$ corresponds to the second hydrogen atom on the water

molecule (H_{W2}), that is not directly involved in the hydrogen bond correlating with the carboxyl atom within a constrained region due to the restrictions on the water molecules' geometry, as discussed above.

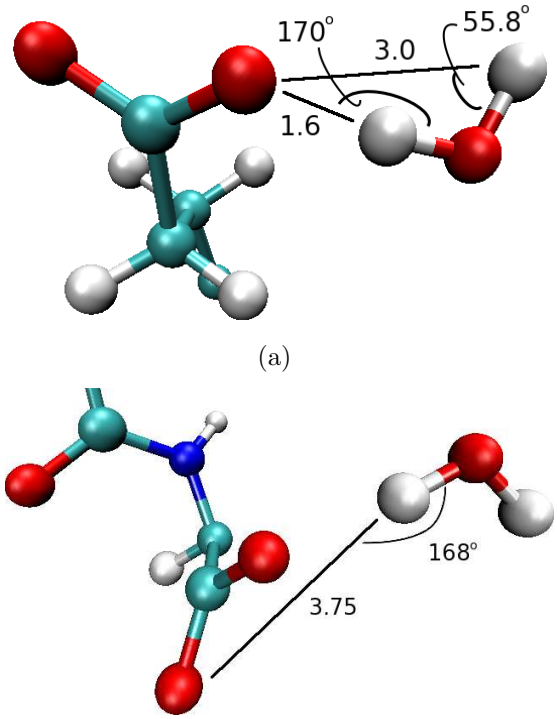


(a) $O_{CE}\cdot\cdot H_W - O_W$ (b) $O_C\cdot\cdot H_W - O_W$

Figure 5.20: 2D radial-angular distribution functions for the GPE-GPE interactions between water and the carboxyl groups in the 1:1:29 Open D system. This correlates the $r_{O_{CE}-H_W}/r_{O_C-H_W}$ distance with the $\theta_{O_{CE}-H_W-O_W}/\theta_{O_C-H_W-O_W}$ angle.

The third peak at $r_{O_{CE}/O_C-H_W} \simeq 3.8$Å, $\theta_{O_{CE}/O_C-H_W-O_W} \simeq 180°$ is at the same radius as the shoulder present in the $O_{CE}/O_C - H_W$ RDFs in figure 5.19. This supports our suggestion that the shoulder is due to a linear relationship between O_{CE}/O_C and H_W atoms that are not directly involved in the hydrogen bond. In figure 5.21 we show a C_α-carboxyl group with a water molecule producing the geometry that is consistent with the shoulder on the $g_{O_C/O_{CE}-H_W}$ RDF. This shows that for a $O_{C1}\cdot\cdot H_{W1} - O_W$ hydrogen bond, the unbonded oxygen in the carboxyl, O_{C2}, and bonded hydrogen from the water molecule, H_{W1}, correlate at ~3.8 Å and $\theta_{O_C-H_W-O_W} > 180°$. A similar motif occurs between the side-chain carboxyl and water molecules, resulting in the shoulder on the trailing edge of the second peak of $g_{H_W-O_{CE}/O_C}(r)$.

Figure 5.22 shows the 2D radial-radial distribution functions correlating the $r_{O_W-O_{CE1}/O_{CE1}}$ and $r_{O_W-O_{C2}/O_{C2}}$ distances in the 1:1:29 Open D system (similar results were obtained for the other systems, and are shown in appendix C, figures C.10 and C.11). We also show the corresponding glutamate-water plots discussed in Chapter 3. As expected from the $O_W - O_{CE}/O_C$ RDFs there is little difference in these 2D radial-radial distributions. Bifurcated motifs exist where one water oxygen atom (OW) is bonded to both carboxyl oxygens, resulting in a region of non-zero intensity at $r_{O_W-O_{CE2}/O_{C2}} \simeq r_{O_W-O_{CE2}/O_{C2}} \simeq 3.2$Å.

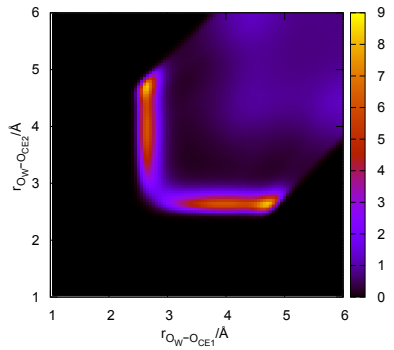


(a)

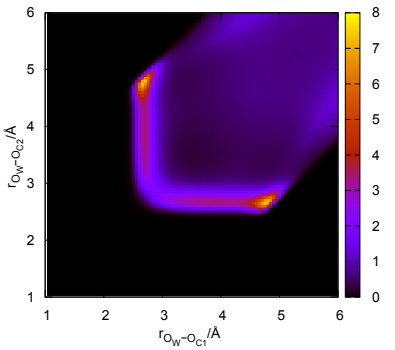
(b)

Figure 5.21: Schematic diagrams showing the three water-carboxyl interactions detected in the 2D radial-angular distribution plots in figure 5.20. Figure 5.21(a) shows the relationship between water and a carboxyl oxygen of a GPE molecule interacting*via*a hydrogen bond of length 1.6 Åand approaching a linear configuration, and the interaction resulting from the correlation of the second hydrogen atom in the water molecule with the same carboxyl-oxgyen at 3.0 Å. Figure 5.21(b) shows therelationship between carboxyl oxygens of the GPE molecule and water molecules resulting in $r_{O_{CE}/O_C-H_W} \sim 3.8$Å, thereby producing a shoulder in the $g(r)_{O_{CE}/O_C-H_W}$ RDF.

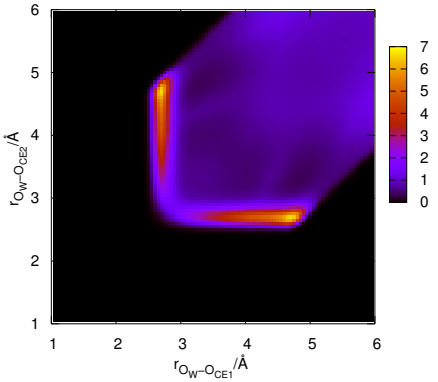
Both carboxyls prefer to form single bonds with water, indicated by the greater intensity in the $r_{O_W-O_{CE1}/O_{C1}} \simeq 3.2$Å, $r_{O_W-O_{CE2}/O_{C2}} > 3.5$Å and $r_{O_W-O_{CE1}/O_{C1}} > 3.5$Å, $r_{O_W-O_{CE2}/O_{C2}} \simeq 3.2$Å regions. We found this reflected in the 2D radial-radial distribution functions correlating $r_{H_W-O_{CE1}/O_{CE1}}$ and $r_{H_W-O_{C2}/O_{C2}}$, shown in appendix C, figures C.12 and C.13, where bifurcated and linear bonds are indicated, but with a preference for single linear bond motifs.



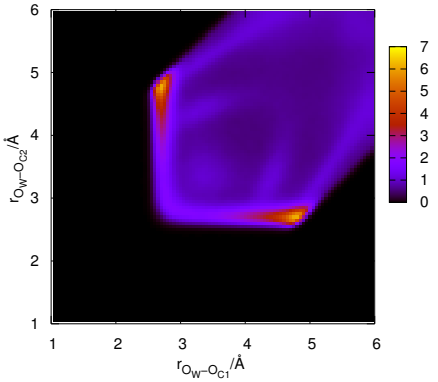
(a) $O_W - O_{CE}$ GPE System 1:29 Open D



(b) $O_W - O_C$ GPE System 1:29 Open D



(c) $O_W - O_{CE}$ Glu



(d) $O_W - O_C$ Glu

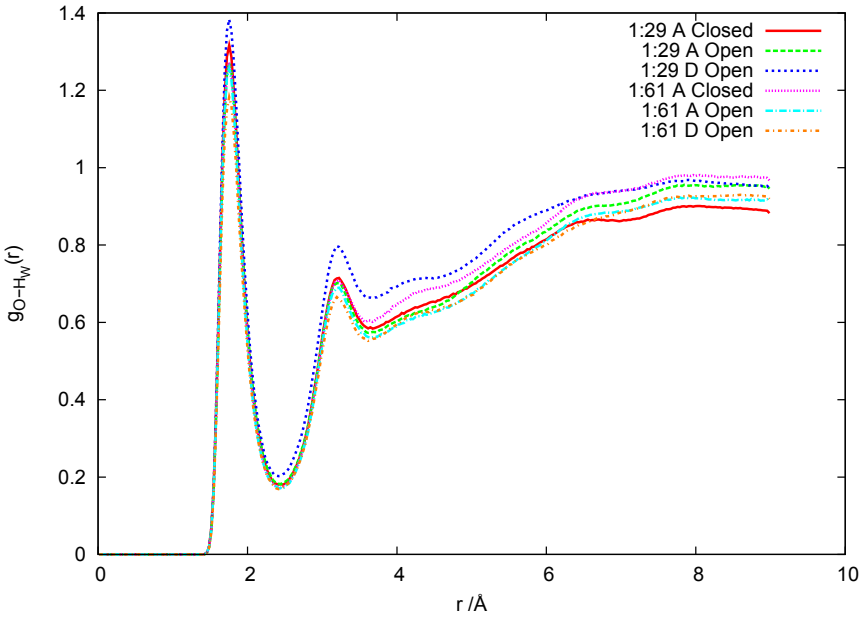
Figure 5.22: 2D radial-radial distribution functions for the GPE-water and glu-water interactions between O_W and O_{CE}/O_C (showing just the 1:1:29 D open system for GPE).

Hydrogen bonds also occur between the backbone oxygen atoms (O) of the GPE molecules and water molecules. Hydrogen bonds also exist between the amine of the GPE molecules and water, as occurred in the glutamate solutions. Figures 5.23 and 5.24 show the RDF plots for the backbone-oxygen and the

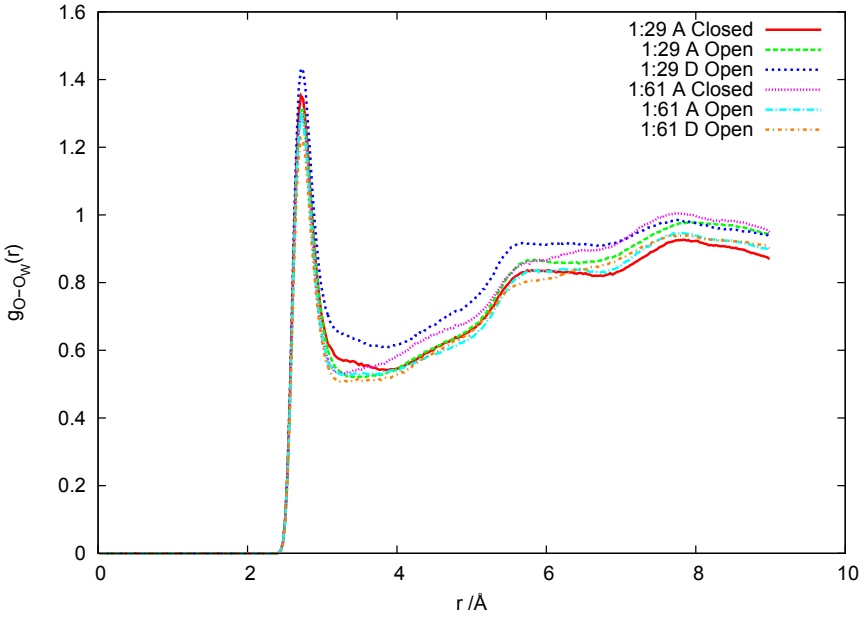
amine interactions with water molecules. There are also very few GPE-GPE bonds formed via the backbone oxygen as the co-ordination number in all six systems is between 1.8 and 1.9. The small co-ordination number and no second peak in $g_{O_W-O}(r)$ indicate that the backbone oxygen only perturbs the water molecules around it minimally. Biologically peptide backbone atoms are often in areas that are generally hydrophobic, which is consistent with this find. This is consistent with the picture obtained from the GPE-GPE bonds, where we found that few bonds formed on the peptide backbone, presumably because of a smaller electric charge on the oxygen atoms as they are part of the peptide bonds and also due to the proximity to positively charged atoms on the peptide backbone.

The amine-water RDFs, figure 5.24, indicate that a second co-ordination shell of water around the amine groups exists, although there was no indication of this in the glutamate solution (shown by the absence of a second peak by the dotted black line in figure 5.24(b)). The amine terminus to the tripeptide is parameterised in a similar fashion to the glutamate molecule, however, in the glutamate molecule the amine is far closer to the negatively charged C_α-carboxyl. We also found that a water molecule could bridge the gap between the C_α-carboxyl and the amine, forming a circle motif, in the glutamate system, which is not possible in the GPE system. Both the proximity of the carboxyl group and the presence of a bridging water molecule may contribute to the absence of a clearly defined second hydration shell around the amine in the glutamate system, and therefore explain the ability of GPE molecules to perturb the water structure enough to form a second shell.

Neither the amine or the backbone oxygen is able to produce bifurcated bonding motifs as shown by the 2D radial-radial distribution plots of 1:1:29 Open D system in figure 5.25. These both show regions of high intensity at $r_{O-H_1} \simeq 1.8$Å, $r_{O-H_2} \simeq 3.1$Å, and $r_{O-H_1} \simeq 3.1$Å, $r_{O-H_2} \simeq 1.8$Å, consistent with a single bond, and zero intensity in the bifurcated bond region at $r_{O-H_1} \simeq r_{O-H_2} \simeq 1.8$Å (where O refers to either O_W or O, and H refers to either H_X or H_W). Similar motifs where found in the 2D radial-radial distribution plots for the other five systems, shown in appendix C, figures C.14 and C.15.

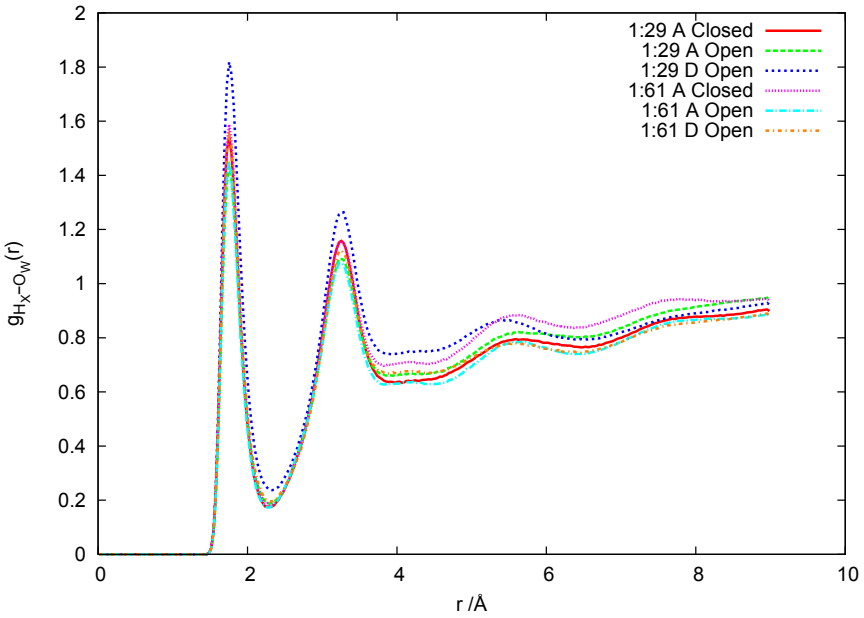


(a) $g_{H_W-O}(r)$

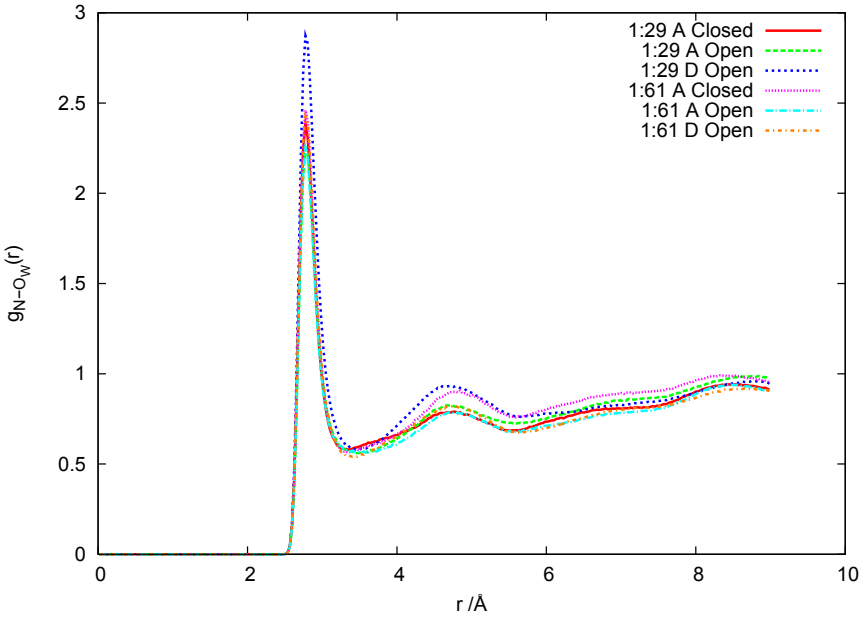


(b) $g_{O_W-O}(r)$

Figure 5.23: Radial distribution functions for the peptide backbone interaction with water.

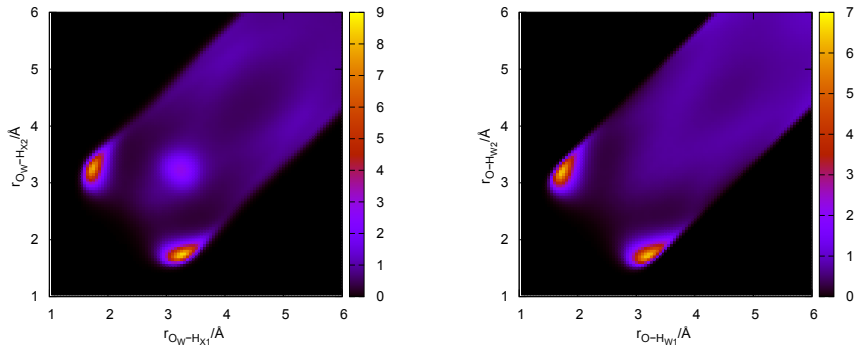


(a) $g_{O_W-H_X}(r)$



(b) $g_{O_W-O}(r)$

Figure 5.24: Radial distribution functions for the GPE amine interaction with water. For comparison we also show the amine-water interaction from the glutamate simulation.



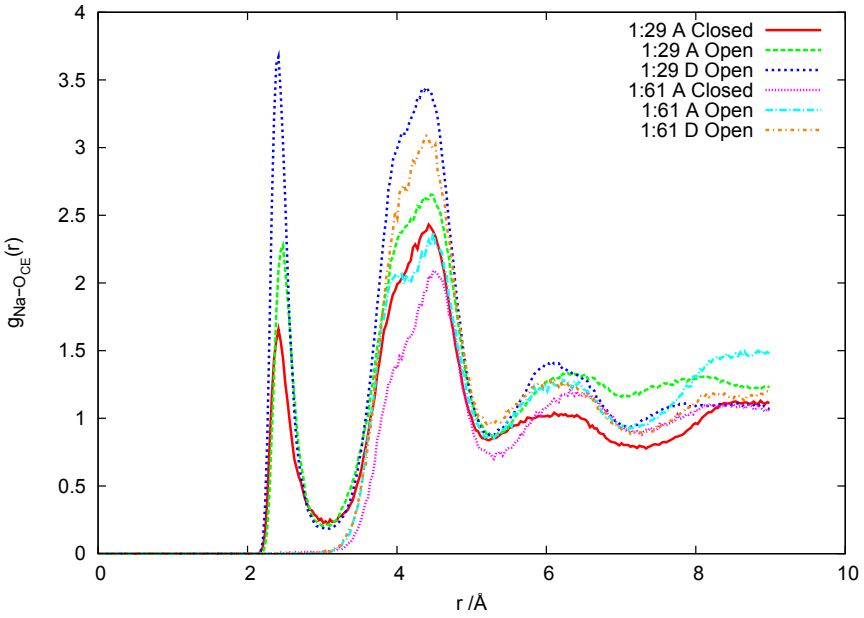
(a) $O_W - H_X$ (b) $O_W - H_W$

Figure 5.25: 2D radial-radial distribution functions for the GPE-water interactions $O_W \cdot\cdot H_X - H_X \cdot\cdot O_W$ and $O \cdot\cdot H_W - H_W \cdot\cdot O$ for the 1:1:29 Open D system.

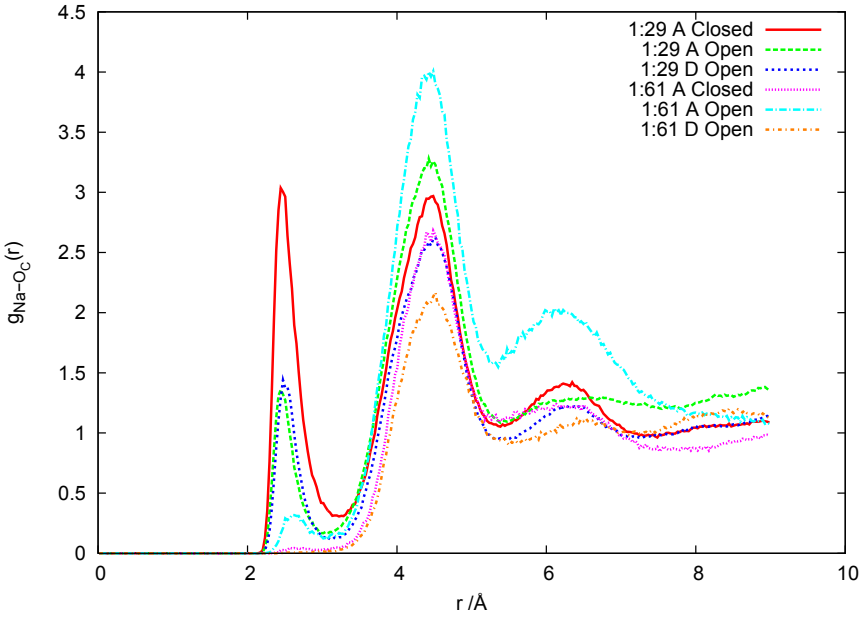
5.4.5 The behaviour of sodium ions in GPE solutions

As with the glutamate solutions, an equal number of GPE molecules and sodium ions are considered in each system to ensure an overall neutral charge. Figures 5.26 and 5.27 shows the radial distribution functions of sodium ions interacting with each of the two carboxyl species, O_{CE} and O_C, the backbone oxygen, O, and the water oxygen, O_W, for the six systems under consideration.

In the carboxyl-sodium RDFs, figure 5.26, there are three clearly defined peaks at $r \simeq 2.5$Å, $r \simeq 4.4$Å and $r \simeq 6.2$Å. For all but the 1:1:29 Closed A system in the $Na^+ - O_C$ RDF and the 1:1:29 Open D system in the $Na^+ - O_{CE}$ RDF, the second peak is of greater intensity than the first. This contrasts with the glutamate solutions where we found a first peak of an order of magnitude larger than the second for both the $Na^+ - O_{CE}$ and the $Na^+ - O_C$ interactions. The first peak in the carboxyl-sodium ion RDFs is due to a sodium ion bonding to a carboxyl-oxygen. The second peak will often be due to the second carboxyl-oxygen interacting with the sodium ions. The greater intensity of the second peak, particularly in the low concentration systems (1:1:61), coupled with the third peak suggests that some of the second peak interaction is due to another atom mediating the carboxyl-sodium ion interaction. This would result in a greater intensity second peak, and the presence of the third peak due to the carboxyl-oxygen not directly involved in the 'mediation'. Indeed, the remarkably low intensity of the low-concentration system RDFs suggests that where possible

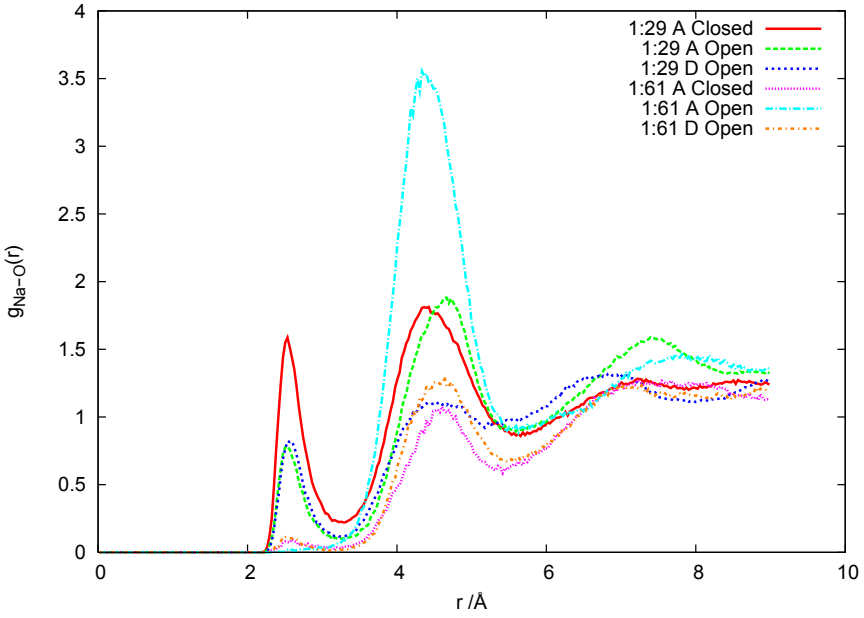


(a) $g_{Na^+-O_{CE}}(r)$

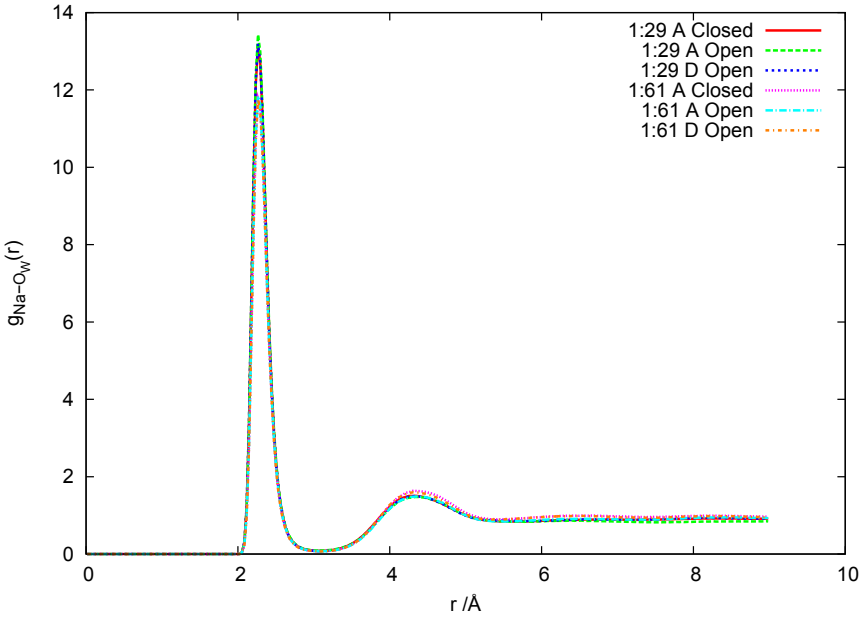


(b) $g_{Na^+-O_C}(r)$

Figure 5.26: Radial distribution functions of the sodium-GPE-carboxyl interactions.



(a) $g_{Na^+-O}(r)$



(b) $g_{Na^+-O_W}(r)$

Figure 5.27: Radial distribution functions for the sodium interactions with the peptide-backbone oxygen, O, of the GPE molecule and water.

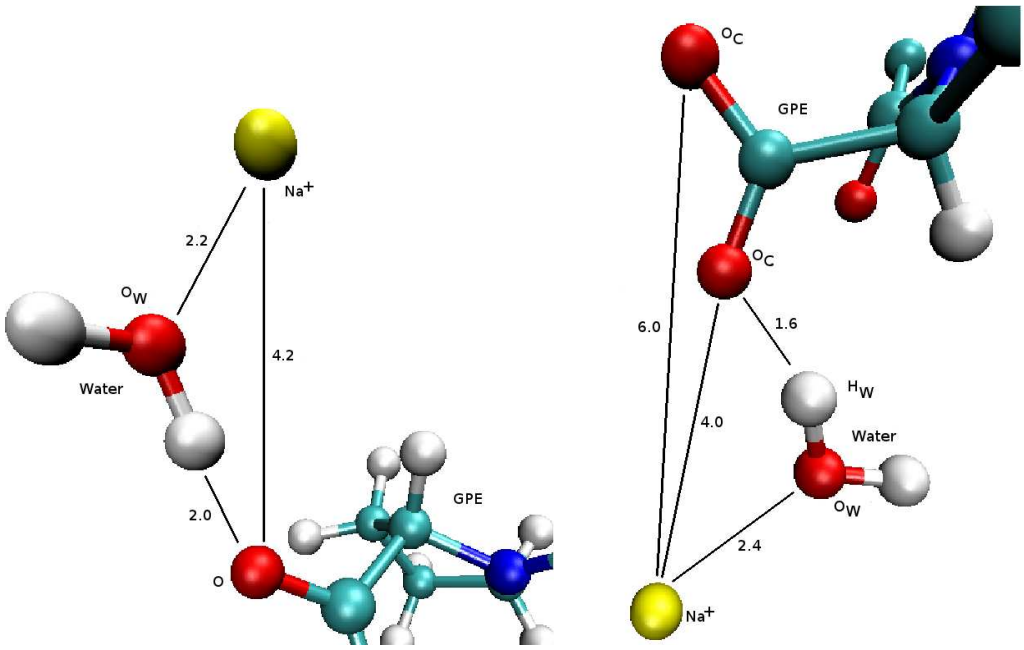
(i.e. at smaller GPE densities), the GPE molecules prefer to not have direct interactions with the sodium ions.

The same trend, although with a less clearly defined third peak, is observed in the sodium-backbone-oxygen RDF (figure 5.27(a)). Here the second and third peak are likely to be due to the second peptide backbone oxygen (one exists on the pro-residue, the second on the gly-residue). The two O-atoms within the same GPE molecule will have greater relative conformational freedom than two carboxyl-oxygens, therefore explaining the different spread of the second and third peaks in this RDF.

In contrast the water-sodium RDF, figure 5.27(b), shows two clear peaks at ∼2.1 Å and ∼4.1 Å. The second peak is broader than the first and indicates a second hydration shell forming around the sodium ions as has been observed in classical simulations of aqueous NaCl solutions by Hermansson *et al.* [51], and also seen in the glutamate solutions studied in chapter 3. The marked decrease in intensities between the first and second peaks indicates that the sodium ion-water interaction is not mediated by another molecule. Indeed, the intensity of the first peak suggests that the water molecules may be the 'mediator' in the sodium ion-GPE interactions.

On inspection of the trajectories we find that water does mediate the interactions between GPE and sodium ions. Schematics of these interactions are shown in figure 5.28. Figure 5.28(a) shows a water molecule forming a bond with both the peptide backbone O atom and sodium ions, producing an $Na^+ - O$ radial distance of $\sim$ 4.2 Å, consistent with the second peaks in the $Na^+ - O$ RDF in figure 5.27(a). Similarly water mediates the interaction between sodium ions and the C_α-carboxyl of the GPE molecule as shown in the schematic in figure 5.28(b). In this case water molecules again form bonds with sodium and with one of the O_C molecules. This produces a $Na^+ - O_C$ radial distance of ∼4.0 Å, consistent with the second peak in the $Na^+ - O_C$ RDFs, and also a radial distance between the O_C atom, not bound to the water molecule, and the sodium ion of ∼6.0 Å, consistent with the third peak of the $Na^+ - O_C$ RDF in figure 5.26(b). A similar motif exists between the side-chain carboxyl and sodium ions producing the second and third peaks of the $Na^+ - O_{CE}$ RDFs in figure 5.26(a).

The 2D radial-radial distribution functions between sodium ions and the

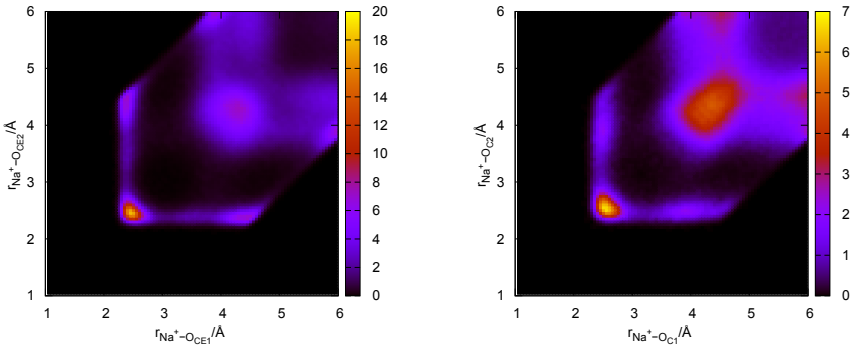


(a) Sodium-Water-O (b) Sodium-Water-OC-OC

Figure 5.28: Schematics showing water mediating the interaction between GPE oxygens and sodium ions.

carboxyl groups, correlating $r_{Na^+-O_{CE1}/O_{C1}}$ and $r_{Na^+-O_{CE1}/O_{C2}}$, also show peaks at $r \simeq 4.1$ Å and $r \simeq 6.0$ Å (figure 5.29). The plots also illustrate a clear preference for direct sodium-carboxyl-oxygen interactions to be*via*the bifurcated state, as indicated by the area of high intensity at $r_{Na^+-O_{CE1}/O_{C1}} \simeq r_{Na^+-O_{CE1}/O_{C2}} \simeq 2.2$Å. The region of greater intensity at $r_{Na^+-O_{CE1}/O_{C1}} \simeq 2.2$ Å, $r_{Na^+-O_{CE1}/O_{C2}} \simeq 4.4$ Å is consistent with the formation of a single $Na^+ - O_{CE}/O_C$ bond, and then the second oxygen in the carboxyl co-ordinating with the sodium ion. The region of high intensity at $r_{Na^+-O_{CE1}/O_{C1}} \simeq r_{Na^+-O_{CE1}/O_{C2}} \simeq 4.2$Å correlate with an $Na^+ - O_W$ bond mediating the sodium ions-carboxyl interaction, but with both carboxyl oxygens bonding to one of the H_W atoms in the mediating water molecule, resulting in both oxygens in the carboxyl co-ordinating at a radial distance of ~ 4.2 Å to the sodium ion as illustrated in the schematic in figure 5.30.

Finally we consider the proportion of sodium ions bonded to between 1 and 6 water and GPE neighbours in figure 5.31. This shows that the majority of sodium ions prefer to form no direct bonds with GPE molecules, and instead have the full complement of six water neighbours. This contrasts with the glutamate



(a) $Na^+ - O_{CE}$ 2D rad rad (b) $Na^+ - O_{CE}$ 2D rad rad

Figure 5.29: 2D radial-radial distribution functions for the GPE-GPE interactions the two species of carboxyl oxygen, O_{CE} and O_C with the Na^+ (sodium ions) for the 1:1:29 open D system. The corresponding figures for the other 5 systems are shown in appendix C, figures C.16 and C.17.

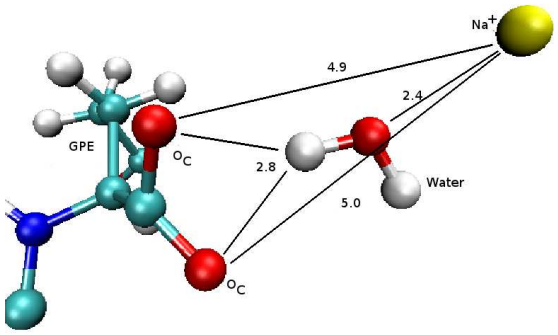


Figure 5.30: Schematic showing water forming bifurcated bonds with the C_α-carboxyl oxygens, and bonding with a sodium ion.

solution where we found that the majority of sodium ions had 4 water and two glutamate neighbours. The reason for the preference to mediate GPE-sodium ion interactions with water molecules, when this did not occur in the glutamate solution is not clear, but is an important artefact in characterising the structure of the GPE system. Clearly there is the possibility for sodium ions to bond with GPE molecules as the charged carboxyl groups should provide an appropriate place for such bonds to occur as in the glutamate solutions. Whether this is due to the GPE molecule preferring to form GPE-water bonds or if it is due to other conditions in the system is not clear. Although it appears it is not due to the concentration used as the same results occur in both the 1:1:29 and 1:1:61 solutions.

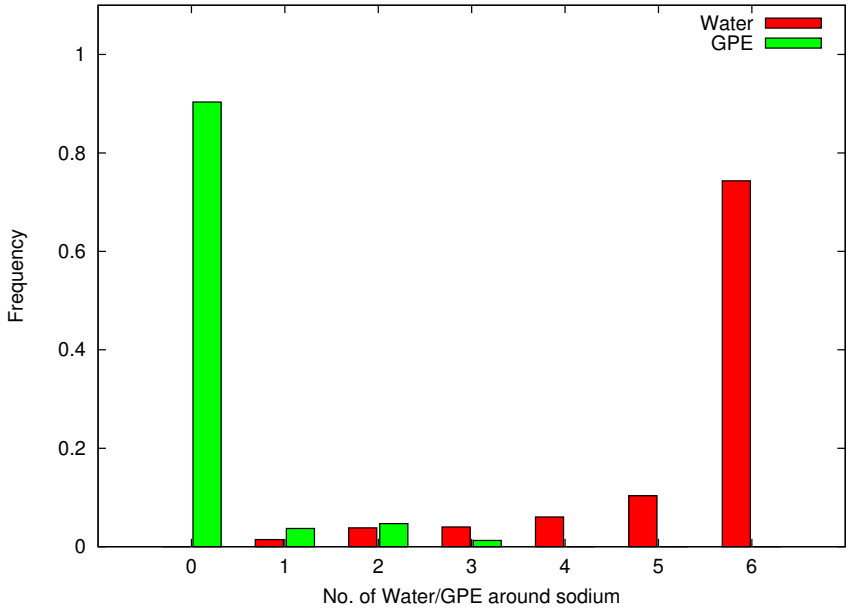
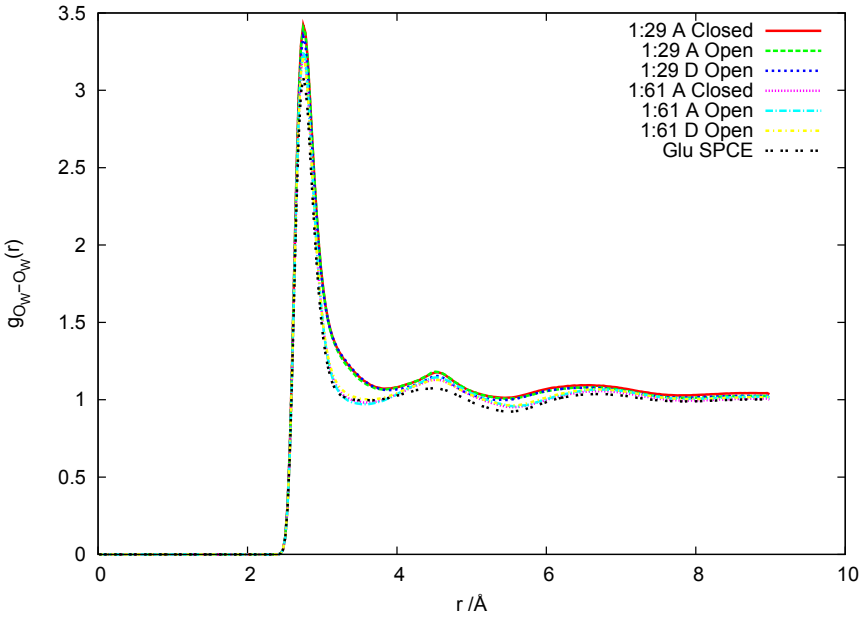


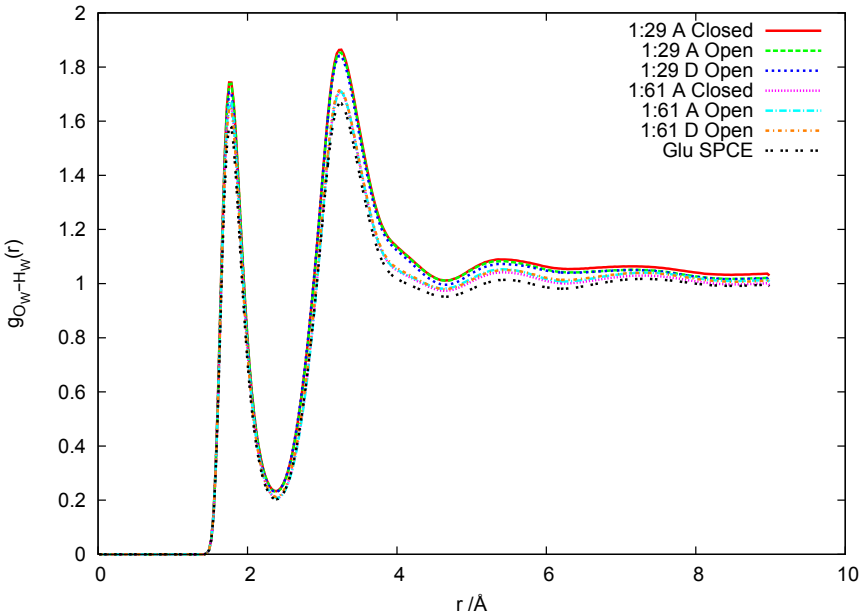
Figure 5.31: Histograms showing the proportion of sodium ions with 1,2...6 water(red) and GPE (green) neighbours as a fraction of the total number of sodium ions in the system for the 1:1:29 Open D system. The corresponding histograms for the other five systems is shown in appendix C, figure C.18.

5.4.6 Water-water interactions in GPE solutions

The radial distribution functions correlating the water oxygen and hydrogen atoms are shown in figure 5.32. The oxygen-oxygen RDF ($g(r_{O_W-O_W})$ in figure 5.32(a)) clearly shows the same peaked structure for all the systems,



(a) $g_{O_W-O_W}(r)$



(b) $g_{O_W-H_W}(r)$

Figure 5.32: Radial distribution functions for the water-water interactions in 1:1:29 and 1:1:61 solutions of GPE.

including the water-water RDFs from the glutamate solution and the pure water SPC/E system discussed in chapter3. A peaks correspond to the first hydration shell at 2.7Å, and a second coordination shell at 4.5Å. We also see that the depth of the first trough at 3.6Å is greatest in the pure water system. The lowest concentration GPE systems (1:1:61) have a similar depth trough to the glutamate solution, and the highest concentration systems (1:1:29) have a poorly defined trough at 3.8Å, therefore indicating that the greater the GPE concentration in the system, the greater the perturbation effect on the bulk water structure.

In figure 5.33 we see that the 2d-radial-radial distribution function indicates that no bifurcated bonds exist between water molecules as there is zero intensity in the region of $r_{O_W-H_{W1}} \simeq r_{O_W-H_{W2}} \simeq 1.8$ Å. This is as we would expect from our results of the pure water in the Glutamate solutions described in chapter 3.

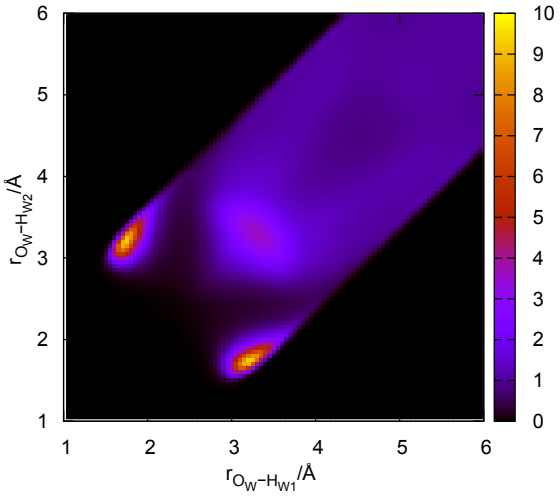


Figure 5.33: 2D radial-radial distribution functions for the GPE-GPE solution $O_W - H_W$ interactions in the 1:1:29 open A system. The other plots for the other systems can be found in figure C.19.

5.5 Conclusions

These results indicate that although there are similarities between the aqueous glutamate and GPE systems, there are also key differences. Firstly, we find that

sodium ions prefer to bond only to water molecules, a stark contrast from the glutamate system. This does not appear to effect the overall perturbation of the bulk water structure as the 1:1:61 system exhibits similar water-structure to the glutamate simulation, although the greater concentration, 1:1:29, systems have a greater impact on water. Secondly, as expected, we see no 'circle'-motif in the GPE system, as the amine and carboxyl groups are too far apart.

We also found that, despite initial configurations favouring either closed or open GPE motifs, all six simulations had a proportion of both closed and open GPE molecules. At both concentrations the Open D configuration had the largest proportion of open molecules. We found that the presence of 'closed' molecules forming intra-molecular bonds between one of the carboxyl groups and the amine, increased the proportion of inter-molecular bonds forming *via* the same carboxyl.

Further consideration of this system is required. In particular, quantum-mechanical work, such as a CPAIMD simulation, would enable verification of some of the motifs found in this classical work, such as the presence of the water molecule mediating the GPE-water interaction. In a similar fashion experimental data would enable verification of the method used for simulation.

We have found that although the GPE tripeptide has the two carboxyl groups and the amine in common with glutamate, that it behaves quite differently in solution with water and sodium ions. Bifurcated bonds between the molecules are still common, but the tripeptide's lack of interaction with sodium ions is somewhat unexpected. This warrants further study; it may be prudent to consider the gly-pro components of the system to see if they are influencing the sodium ions and thereby reducing the GPE-sodium ion interactions .

As GPE has been found to have neuroprotective effects during brain hypoxia, including when administered *via* injection, implying that it is able to cross the blood-brain membrane, these results may have an impact on understanding how GPE provides protection [52]. In addition GPE may provide treatment to neuro-degenerative disorders such as Parkinsons, although the mechanism for this is currently unknown [53]. In both instances the mechanism for this protection is not known, although it requires transition through of the GPE fragment through cell membranes. Further studies of GPE in solution with other amino acids could aid the understanding of both of these mechanisms. For instance, if GPE and glutamate bind together through long-lived hydrogen bonds this would suggest

that GPE effects the presence of excessive glutamate in the extra-cellular space that is present during hypoxia. As there is no 'circle-motif' similar to that found in the glutamate system, further studies may be useful to assess if the presence of GPE stops the stable circle motifs forming, allowing the glutamate molecules to move around more freely, and thereby bind to sodium ions, crossing the membrane barrier more quickly.

Unlike the glutamate solutions, the sodium ions preferred to form a hydration shell of six water molecules at both concentrations of GPE:sodium ions:water. The sodium ions also preferred to avoid bonds with GPE molecules, which contrasts with neighbours of 2,3 and 4 glutamate molecules around sodium ion discussed in chapter 3. This may have a direct impact on the biological mechanisms which cause GPE to provide neuro-protection. As glutamate holds onto sodium ions, the GPE molecules release the ions, perhaps allowing them to provide the membrane-pump action required to remove the excess neurotransmitters from around synapses. Again this would suggest that simulations of GPE and glutamate in solution would provide valuable information in understanding the biological mechanisms associated with GPE and glutamate.

Chapter 6

Conclusion

The work examines the structural properties of two biomolecular systems, aqueous L-glutamate (glu) and aqueous L-glycine-L-proline-L-glutamate (GPE). In each case, they were modelled *via* computer simulation, in solution with water and sodium counter-ions.

Firstly, a classical simulation of aqueous glu enabled the characterisation of the glu-glu interactions, using the CHARMM22 force-field and SPC/E, TIP3P and F3C water potentials. We found that the two charged carboxylate groups of the glu molecules behave differently in interactions with not just water, as previously identified by Leenders *et al.* [44] and Daub *et al.* [45], but also when interacting with other glu molecules. In particular, the C_α-carboxyl of the glu molecule is able to form a 'circle'-motif where a water molecule is simultaneously co-ordinated with both the carboxyl and the amine of the same glu molecule. This results in a shoulder between the first and second peak of the water-glu radial distribution function, which is not present in the side-chain carboxyl distribution, due to the stabilisation of the water molecule at a distance that is not fully consistent with a strict hydrogen-bond interaction.

The classical simulations of glu also show that glu prefers to bond to the sodium ions present in the solution *via* the side-chain rather than the C_α carboxyl. There are also a greater number of water-glu bonds formed *via* the C_α-carboxyl group than with the side-chain carboxyl, presumably due to the presence of the sodium ions around the side-chain. The Car-Parrinello *ab initio* Molecular Dynamics (CPAIMD) simulations also demonstrate the same trend although, due to the small system size giving poor statistics the results are less conclusive.

However, as the CPAIMD simulation was initiated from the fully equilibrated classical simulation configuration, we suggest that this preference for sodium ions to interact with the side-chain in both sets of simulations may be a result of the initial configuration and slow dynamical movement in the system limiting the ability of the sodium ions to co-ordinate with other groups.

The apparent slow dynamics limit the efficacy of the simulations as a method of fully sampling the configuration landscape of the system. Thus the dynamics of the glu simulations were studied*via* consideration of the bond lifetimes from auto-correlation functions and self-diffusion coefficients. The sodium-glu interactions were found to be the longest lived bonds within the classical simulation. The glu-water and glu-sodium bond lifetimes were also greater than the water-water bonds in a pure water simulation. This suggests that the energetic barriers between conformations in the glu system are sufficiently high that classical simulations do not provide effective sampling of the conformational landscape. Thus the enhanced sampling technique, parallel tempering was applied.

The use of parallel tempering simulations enabled us to obtain different sizes of clusters than had been available from the classical simulations. Preference for the type of glu-glu bonds, whether*via* the side-chain or the C_α-carboxyl, were also affected by the parallel tempering, indicating a strong dependence on initial starting configuration for the type of interaction that was observed. The propensity of the 'circular'-motif was also reduced in the parallel tempering simulations. The presence of the sodium ion appears to affect the ability of a water molecule to form a bridge between the amine and C_α-carboxyl of a glu-molecule, and thus the use of a different potential for the sodium ion in the parallel tempering simulations may explain the reduction in the proportion of circle motifs being formed.

The parallel tempering simulations also indicated a reduction of bifurcated bonding motifs, particularly*via* the C_α-carboxyl compared to the classical and CPAIMD simulations of the aqueous glu system. There is a general consensus that empirical force-fields over-estimate bifurcated motifs in classical simulations. However, the presence of bifurcated motifs in the CPAIMD study, and the reduction of them in the classically simulated parallel tempering work suggests that in this instance the presence or lack of bifurcation, suggests that bifurcated bonding motifs exist in both the glutamate and GPE systems. This area needs

more investigation, using simulations of other aqueous solutions, before definitive conclusions about the effect of classical force-fields on bifurcation can be drawn.

Finally, a classical simulation of the GPE tripeptide in solution with water and sodium ions was considered. Six systems were studied, initiated using three different initial conformations of the GPE molecule, each simulated at two different GPE:sodium:water concentrations. Here we found that GPE-GPE bonds preferred to form*via* the side-chain carboxyl rather than the C_α-carboxyl in all six simulations, irrespective of the initial GPE-backbone conformation. In addition, despite the six different systems, all the simulations showed GPE molecules forming intra-molecular bonds where the side-chain of the glu residue of the tripeptide co-ordinated with the N-terminal amine group of the gly residue, although the proportion of molecules forming the intra-molecular bond is greatest in the simulations where the initial configuration of the GPE molecule favoured this type of bond.

We also find that the amine group of GPE molecules can form a distinct second-hydration shell, however, this level of perturbation of the water structure is not observed in the glu simulations. It is suggested that this difference is due to the inability of GPE molecules to form the glu 'circle'-motif, as well as the closer proximity of the carboxyl groups to the amine in a glu molecule compared to GPE. The distance between the C_α-carboxyl and the amine groups in the GPE molecules prohibits the bridging of the two groups by a single water molecule. The presence of a water molecule bridging the gap between the groups disturbs the hydration shell around the amine, and therefore may reduce the structuring of the water in this region. In addition, in GPE molecules, the distance between the positively charged amine and the negatively charged carboxyls is far greater, thereby increasing the overall impact the amine may exert on the surrounding water molecules.

The main aim of this body of work is to provide a thorough understanding of glu and GPE solutions from computer simulations. However, it has also highlighted some of the technical challenges associated with computer simulation. Firstly, although first-principles, quantum-mechanical simulations are the ideal solution to computer simulation techniques, in that they should ideally exactly reproduce the physical interactions of a system, this approach has three main limitations; system size, the time-scales available, and the approximations used

to make the calculations more tractable, such as the use of the local-density approximation and exchange correlation functionals. Although computational resources are growing rapidly, the limitations are still significant. This study in particular highlights the inability of quantum-mechanical simulations to study these systems on time scales long enough to provide meaningful results that do not depend on the original structure. However, the time-scales available to classical simulations are also not necessarily sufficient, as highlighted by the dependence of the results on the original structure in the glu simulations. The small system size of quantum mechanical simulations means that for amino acid simulations sufficient data can not be obtained for meaningful statistics of solute-solute interactions. This does not mean that the results obtained from the classical and Car-Parrinello simulations are not useful, however, they should not be considered in isolation.

Classical, empirical force-field calculations have their own set of problems. They are by definition an approximation of the real system, and their success depends greatly on the accuracy of the force-field for the characteristics under consideration. Often a force-field is constructed to imitate particular properties of a system for a particular set of conditions, and therefore may not easily be transferred to another system. However, empirical force-fields do enable larger system sizes and greater time-scales than quantum mechanical approaches. Classical Molecular Dynamics simulations, particularly of systems with the complex energetic landscapes prevalent in biomolecular systems, are often not simulated for a long enough time to overcome many of the energetic barriers between molecular conformations.

Currently, to our knowledge, parallel tempering has only been used in classical simulation and has not been implemented using an *ab initio* approach, and thus is limited in efficacy by the limitations of empirical force-fields. Parallel tempering is also hindered by the problems inherent in choosing an appropriate temperature set, as this can affect the efficacy of the simulations. Thus careful consideration should be given during the set-up of such simulations. The approach used in this work chooses the upper temperature after consideration of the results of short MD simulations at a variety of different temperatures, and the configurations that this enabled the system to sample. An initial temperature set was then devised to span the temperature range using a geometric progression through temperature

space, with further modifications to the temperature set being used to iteratively optimise the probability of temperer exchange.

Full analysis of all 32 temperers from the parallel tempering simulation was judged unnecessary because it was unlikely that significantly different information would be obtained from each of them. As consideration of just the lowest temperature system would not provide a complete picture of the configurations occurring in the simulation, the distribution of glu clusters sizes was considered. From this five different systems were identified as representative of the entire set of 32 temperers, and these were then subjected to further simulations.

In an analogous situation, there were fifteen different initial configurations of the GPE molecule available from the protein data bank that could have been used in the simulation of aqueous GPE solutions. Simulating all fifteen conformations would have required large computational resources and may have resulted in a very complex set of results for analysis that would probably not offer fifteen sets of significantly different data. A systematic investigation of the different motifs available enabled selection of just three different configurations to be considered for simulation. Conformational comparison of Ramachandran maps provided a repeatable and scalable approach to this selection method. This technique could be applied to systems with far more configurations, and enables the reduction of a large set of configurations into a set of realistic size for computation.

The parallel tempering and the GPE simulations highlight the importance of designing a simulation protocol. Often the main consideration in setting up simulations is reproducing important structural features, and converging energies. However, as shown here, the most useful information can be obtained from simulations that consider aspects of the problem before and after the simulation itself, which in turn may provide more useful material for analysis. In particular, utilising previously obtained structures, as in the GPE case, or structures obtained from parallel tempering that might not otherwise be considered, can enhance the ability of Molecular Dynamics to obtain information not available *via* other means.

In summary, we consider it important to understand the limitations of any simulation technique. In a similar fashion to never considering simulation without also considering experimental data to verify your initial methodology, single computational methods should not necessarily be considered in isolation. The

best information can be obtained from the triangulation of data from a variety of approaches. We have provided two interesting approaches to molecular simulation analysis. Firstly, the analysis of the parallel tempering simulations are not just concerned with the final configuration of the bottom temperature system, or with one individual temperer, but instead consider all 32 temperers, before identifying five that are then subjected to further classical simulations. Secondly, using different conformations of the GPE molecule to initiate different simulations has enabled better sampling of the aqueous GPE solutions than if just one conformation had been chosen. This is a quick method for overcoming some of the energetic impediments that effect classical simulations, requiring fewer computational resources than parallel tempering, although it is not necessarily as effective as enhanced sampling techniques.

6.1 Areas of further interest

This work highlights the strengths and weaknesses of using empirical models for the simulation of biomolecules. The multitude of potentials available, particularly for water molecules and ions, may give different results. The effect of the sodium ion potential on the 'circle'-motif needs to be investigated further, as although no single potential has been singled out as the best forcefield, if the force-field of sodium is found to significantly effect the behaviour of water molecules interacting with biomolecules when in the presence of sodium ions, this needs to be considered in future biomolecular studies. If the type of ion potential does affect the formation of the 'circle-motif' then questions need to be asked about which potential is the most appropriate in the simulation of this type of system. Due to the prevalence of water and ions in biological systems, understanding the complex solute-solvent-ion interactions may become more and more important as our understanding of biology develops.

Given the importance of GPE to the human body, in particular as it is the terminal tripeptide of the insulin-growth factor-1 (IGF-1), further investigation of GPE would be of great benefit. *Ab initio* calculations would enable verification of the classical results, but a representative system size would require computational resources that are not currently readily available. It is hoped that meaningful calculations of both GPE-GPE and glu-glu interactions using

quantum mechanical methods may be available in the near future. Following on from this we would again encourage the use of advanced sampling techniques for simulations of all bio-molecules, and in particular the GPE solutions studied here.

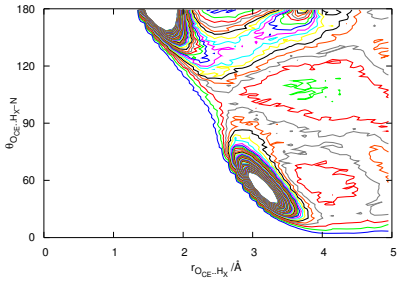
It would be interesting to take this 'bottom-up' approach one step further and simulate the entire IGF-1 polypeptide, although several smaller increments may be required of shorter peptides before the entirety of IGF-1 is considered.

Finally it would be of great interest to obtain experimental data from neutron diffraction on the GPE system to enable comparison and verification with simulation.

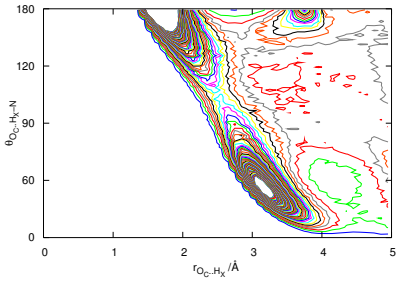
Appendix A

Additional information on the structure of glutamate

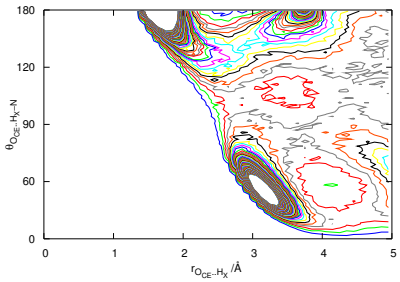
The following plots are supplementary to the discussion of the glutamate structure discussed in chapter 3.



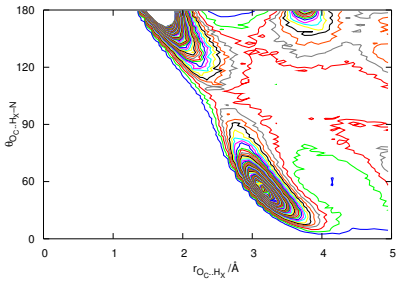
(a) F3C $O_{CE}\cdot\cdot H_X - N$



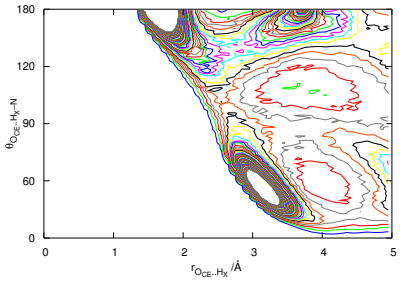
(b) F3C $O_C\cdot\cdot H_X - N$



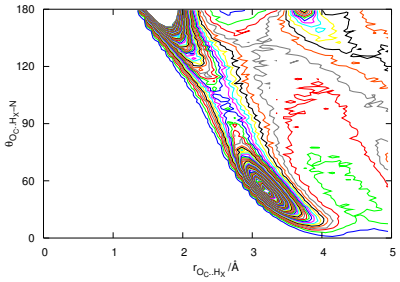
(c) SPCE $O_{CE}\cdot\cdot H_X - N$



(d) SPCE $O_C\cdot\cdot H_X - N$

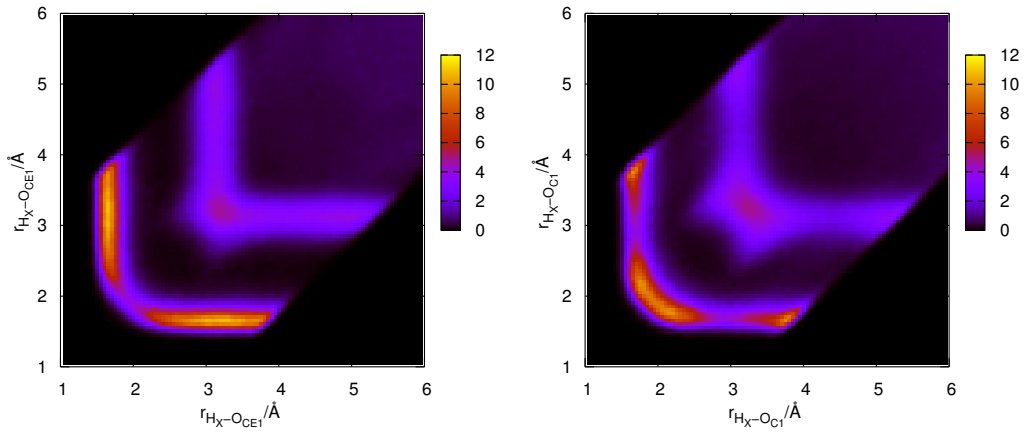


(e) TIP3P $O_{CE}\cdot\cdot H_X - N$



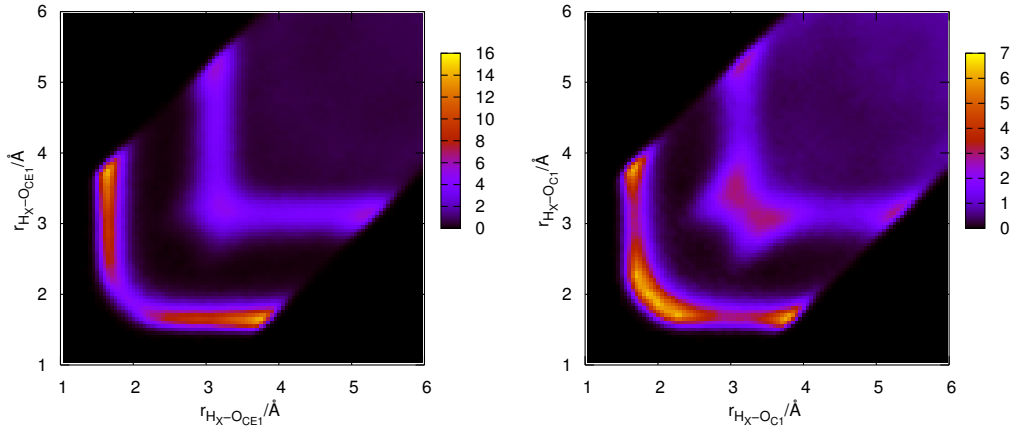
(f) TIP3P $O_C\cdot\cdot H_X - N$

Figure A.1: 2D radial-angular distribution functions for the Glu-Glu interactions $O_{CE}\cdot\cdot H_X - N$ and $O_C\cdot\cdot H_X - N$ in the solutions of CHARMM22 Glutamate with F3C, SPC/E and TIP3P water molecules.



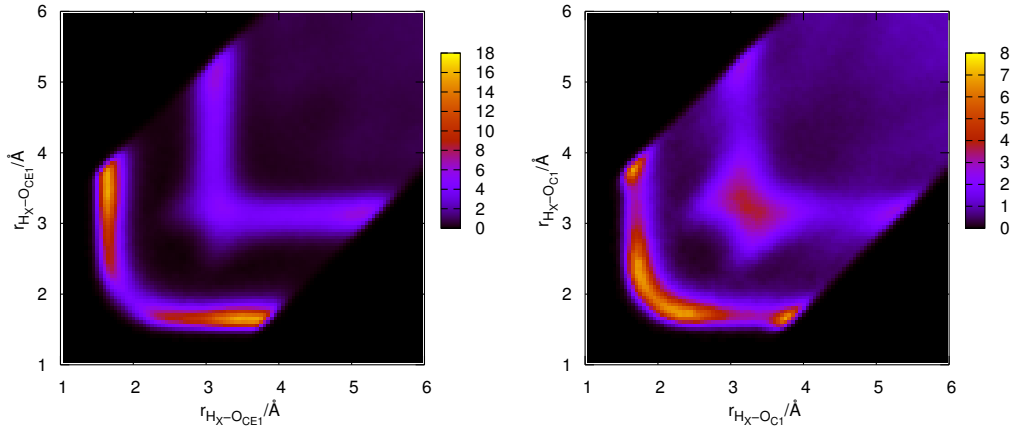
(a) F3C $H_X \cdot\cdot O_{CE}$

(b) F3C $H_X \cdot\cdot O_C$



(c) SPC/E $H_X \cdot\cdot O_{CE}$

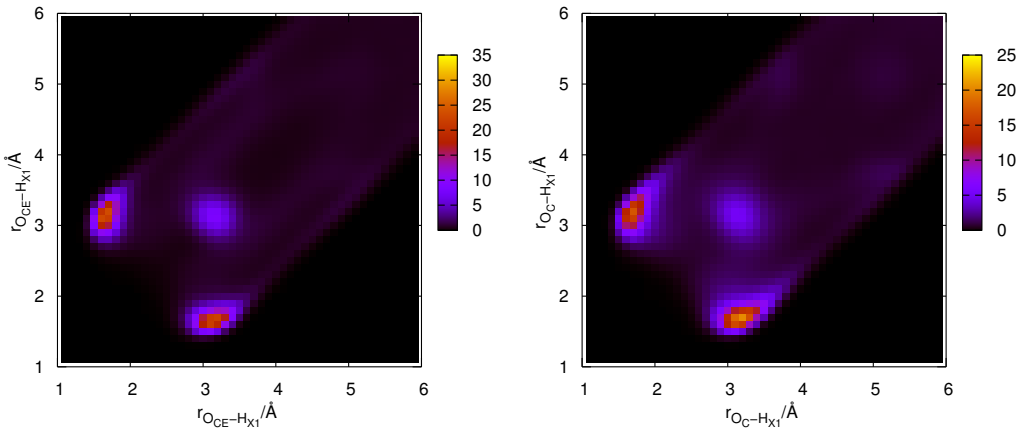
(d) SPC/E $H_X \cdot\cdot O_C$



(e) TIP3P $H_X \cdot\cdot O_{CE}$

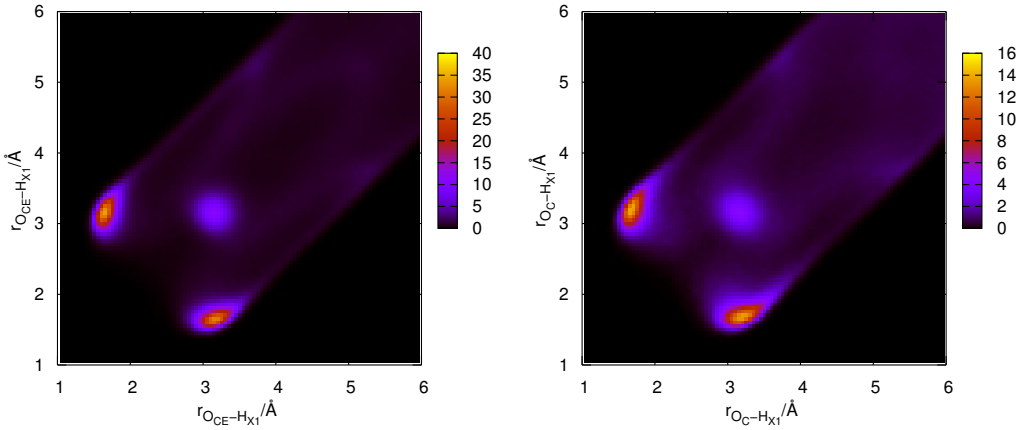
(f) TIP3P $H_X \cdot\cdot O_C$

Figure A.2: 2D radial-radial distribution functions for the Glu-Glu interactions correlating the distance $r_{HX1\cdot\cdot OCE1/OC1}$ with $r_{HX1\cdot\cdot OCE2/OC2}$ in solutions of CHARMM22 Glutamate with F3C, SPC/E and TIP3P water molecules.



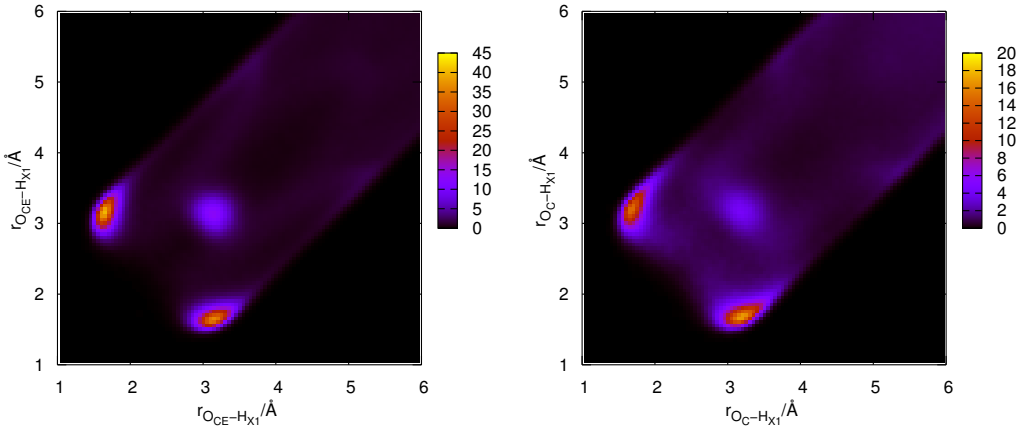
(a) F3C $O_{CE}\cdot\cdot H_X$

(b) F3C $O_C\cdot\cdot H_X$



(c) SPC/E $O_{CE}\cdot\cdot H_X$

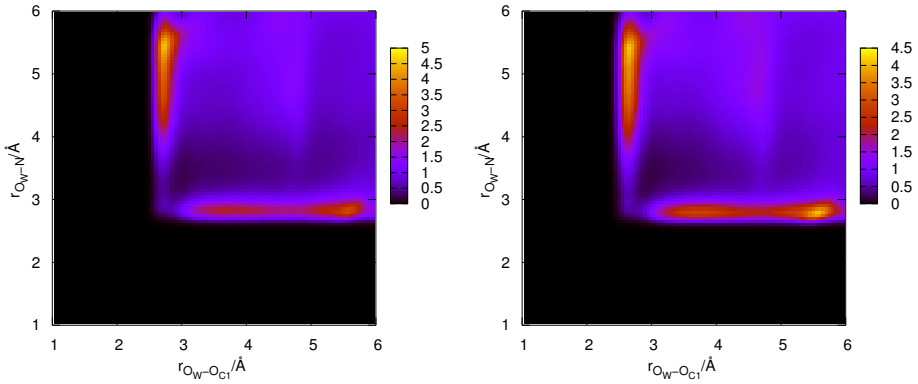
(d) SPC/E $O_C\cdot\cdot H_X$



(e) TIP3P $O_{CE}\cdot\cdot H_X$

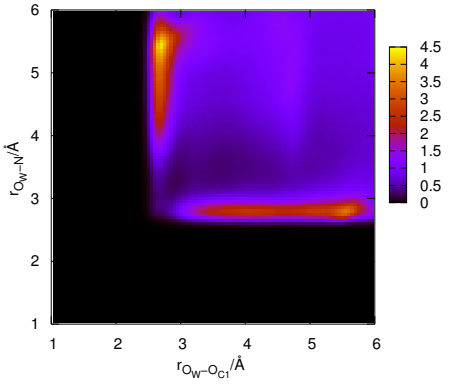
(f) TIP3P $O_C\cdot\cdot H_X$

Figure A.3: 2D radial-radial distribution functions for the Glu-Glu interactions correlating the distance $r_{O_{CE1}/O_{C1}\cdot\cdot H_{X1}}$ with $r_{O_{CE1}/O_{C1}\cdot\cdot H_{X2}}$ in solutions of CHARMM22 Glutamate with F3C, SPC/E and TIP3P water molecules.



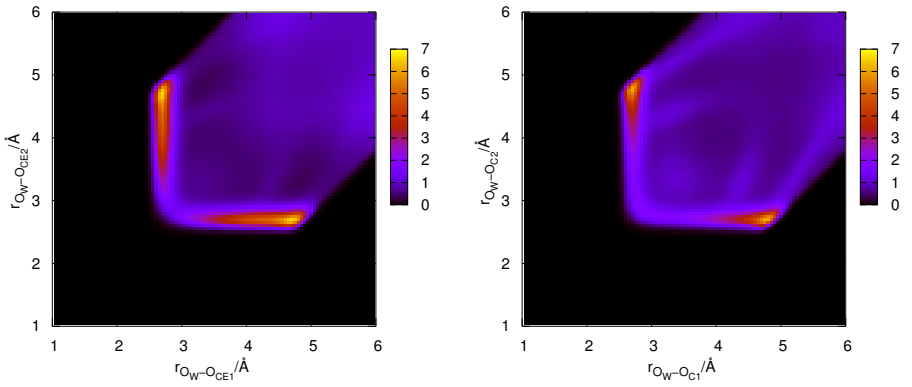
(a) F3C

(b) SPC/E



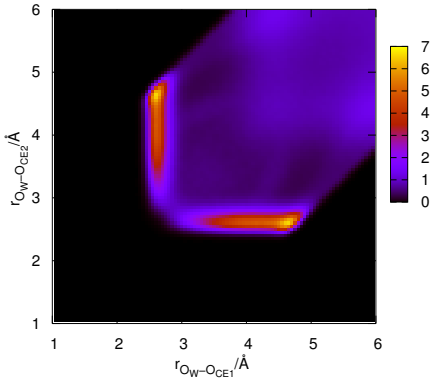
(c) TIP3P

Figure A.4: 2D radial-radial distribution functions for the Glu-Glu interactions correlating the distance $r_{O_{C1}\cdot\cdot O_W}$ with $r_{N\cdot\cdot O_W}$ in solutions of CHARMM22 Glutamate with F3C, SPC/E and TIP3P water molecules.

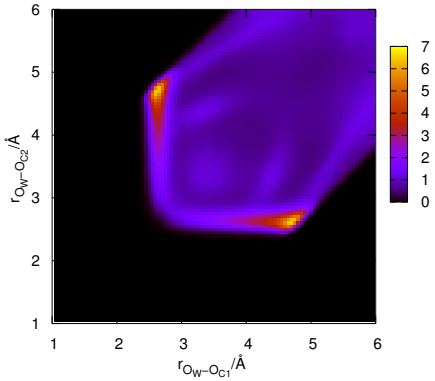


(a) F3C $O_W \cdots O_{CE}$

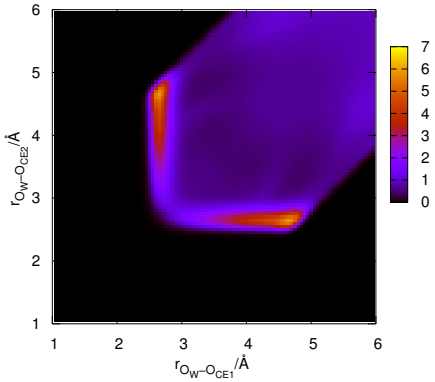
(b) F3C $O_W \cdots O_C$



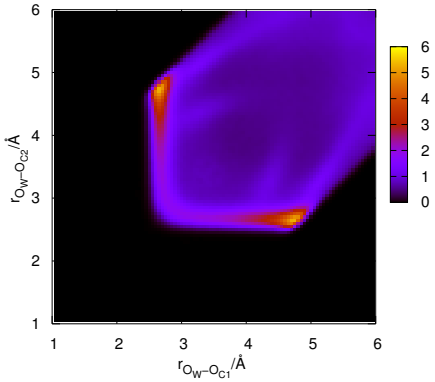
(c) SPCE $O_W \cdots O_{CE}$



(d) SPCE $O_W \cdots O_C$

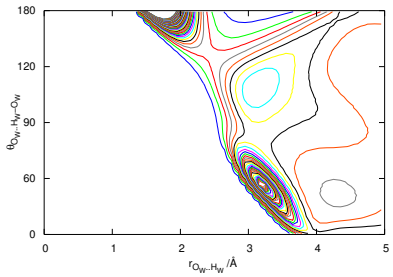


(e) TIP3P $O_W \cdots O_{CE}$

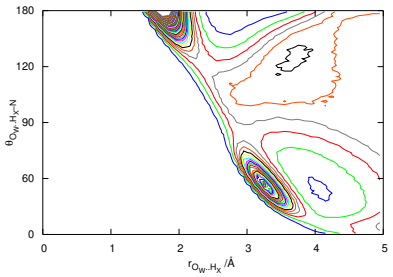


(f) TIP3P $O_W \cdots O_C$

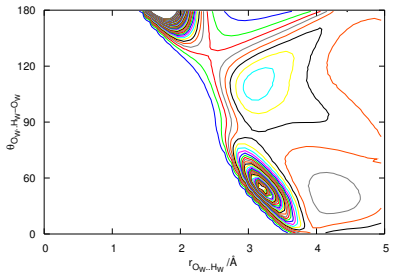
Figure A.5: 2D radial-radial distribution between F3C water-oxygen (O_W) and each of two glu-oxygen species.



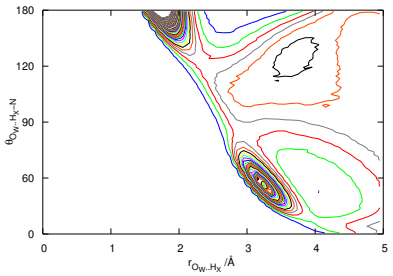
(a) F3C $O_W \cdot\cdot H_W - O_W$



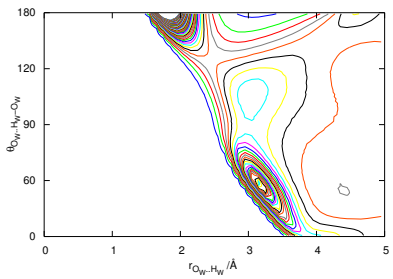
(b) F3C $O_W \cdot\cdot H_X - N$



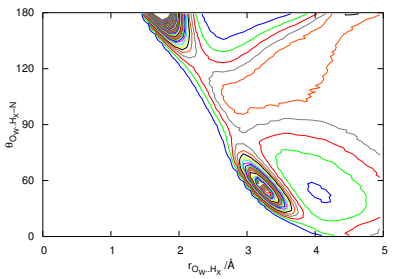
(c) SPCE $O_W \cdot\cdot H_W - O_W$



(d) SPCE $O_W \cdot\cdot H_X - N$

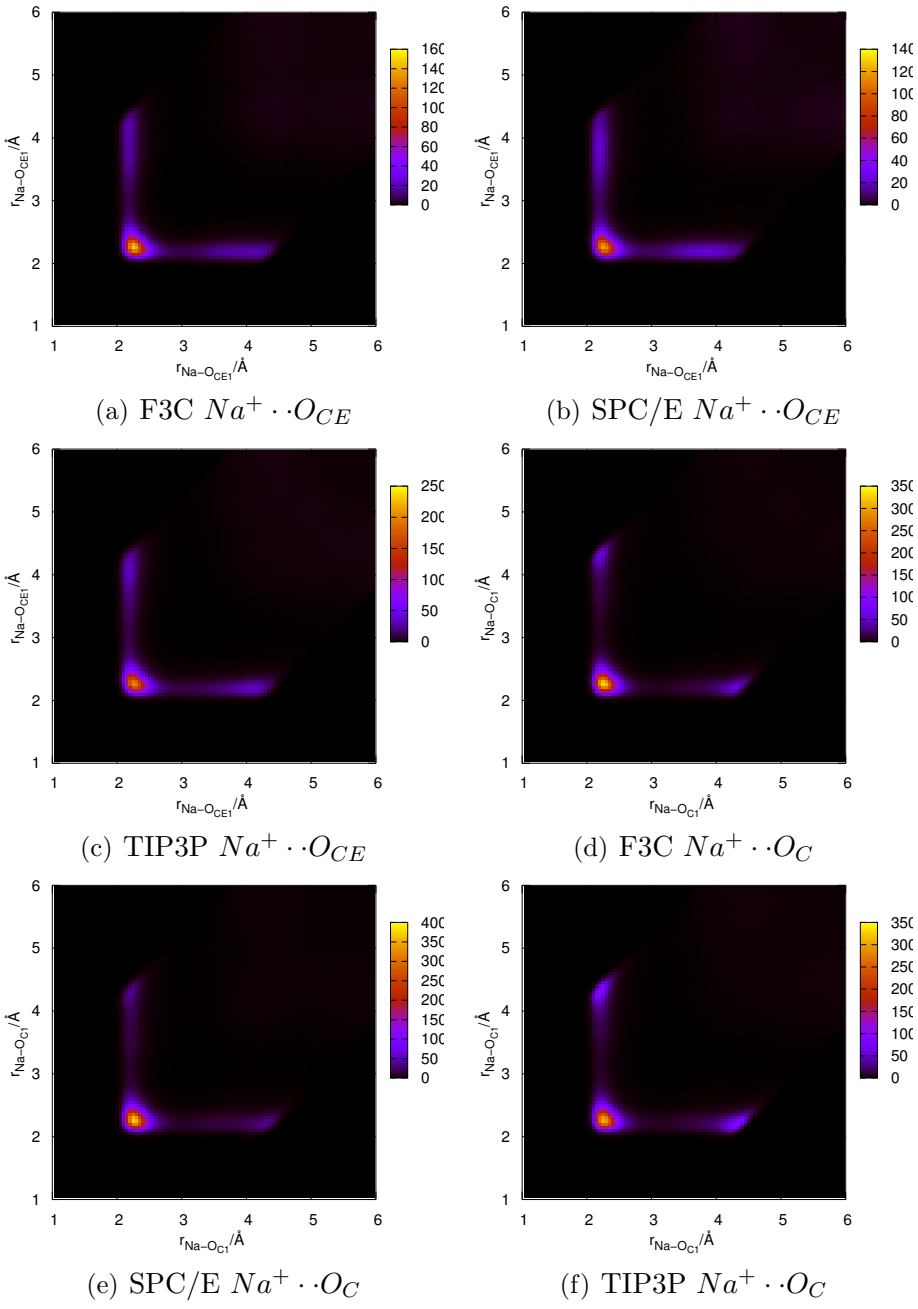


(e) TIP3P $O_W \cdot\cdot H_W - O_W$



(f) TIP3P $O_W \cdot\cdot H_X - N$

Figure A.6: Comparison of the 2D radial-angular distribution for a water-water hydrogen bond, $O_W \cdots H_W - O_W$, and for the glutamate-amine-water hydrogen bond, $O_W \cdots H_X - N$ (from the F3C solution, with similar results obtained for SPC/E and TIP3P).

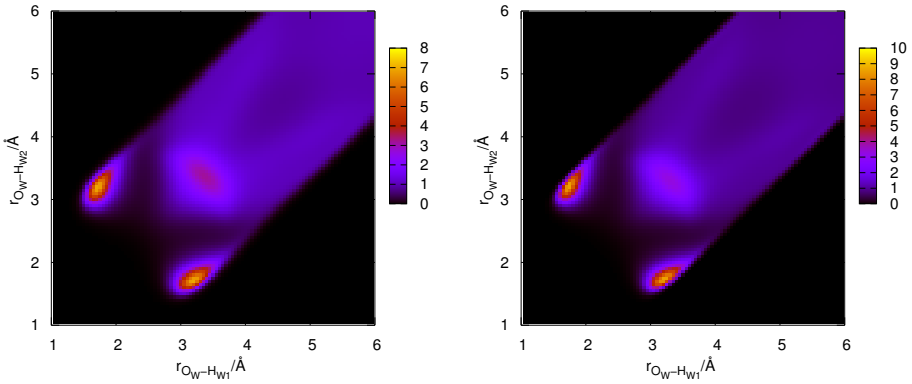


(a) F3C $Na^+ \cdot\cdot O_{CE}$ (b) SPC/E $Na^+ \cdot\cdot O_{CE}$

(c) TIP3P $Na^+ \cdot\cdot O_{CE}$ (d) F3C $Na^+ \cdot\cdot O_C$

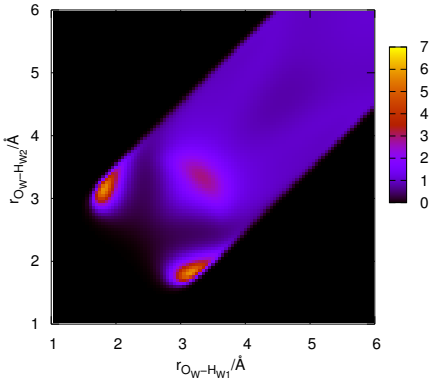
(e) SPC/E $Na^+ \cdot\cdot O_C$ (f) TIP3P $Na^+ \cdot\cdot O_C$

Figure A.7: 2D radial-radial distribution functions for the Glu-Glu interactions correlating the radial distances $r_{Na_1^+ \cdot\cdot O_{CE1}/O_{C1}}$ and $r_{Na_1^+ \cdot\cdot O_{CE2}/O_{C2}}$ in solutions of CHARMM22 Glutamate with F3C, SPC/E and TIP3P water molecules.



(a) F3C

(b) SPC/E



(c) TIP3P

Figure A.8: 2D radial-radial distribution functions for the water-water interactions correlating the radial distances $r_{O_{W1}\cdot\cdot H_{W1}}$ and $r_{O_{W1}\cdot\cdot H_{W2}}$ in solutions of CHARMM22 Glutamate with F3C, SPC/E and TIP3P water molecules.

Appendix B

Additional information on glutamate dynamics and parallel tempering analysis

The following is supplementary material to the discussion in chapter 4.

Donor-Acceptor		Hydrogen-Acceptor	
Type	Cut-off (Å)	Type	Cut-off (Å)
$N - O_C$	3.78	$H_X - O_C$	2.08
$N - O_{CE}$	3.53	$H_X - O_{CE}$	2.43
$N - O_W$	3.48	$H_X - O_W$	2.38
$O_W - O_C$	3.13	$H_W - O_C$	2.43
$O_W - O_{CE}$	3.18	$H_W - O_{CE}$	2.38
$O_W - O_W$	3.73	$H_W - O_W$	2.38
$Na^+ - O_C$	3.13		
$Na^+ - O_{CE}$	3.03		
$Na^+ - O_W$	3.08		

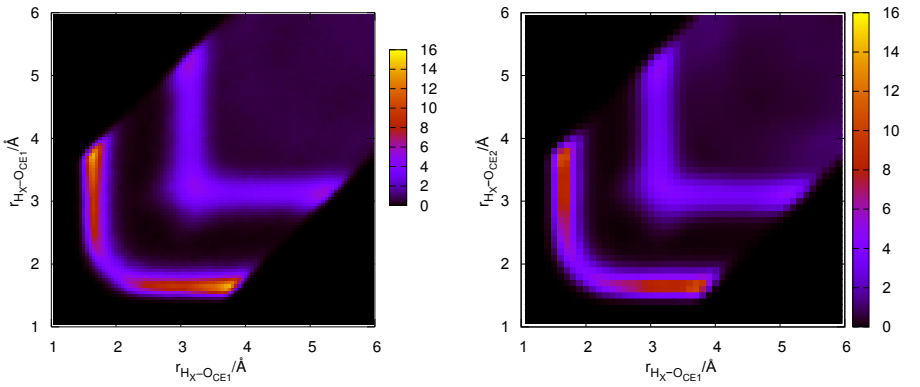
Table B.1: Radial cut-offs used to define a 'hydrogen-bond'.

System	Interaction	Lifetimes τ [ps]			
		t^*=0.0	t^*=0.05	t^*=0.10	t^*=0.50
Glu F3C	$O_{CE} - N$	0.104	0.125	0.242	0.441
Glu SPC/E	$O_{CE} - N$	0.108	0.147	0.297	0.541
Glu TIP3P	$O_{CE} - N$	0.105	0.133	0.250	0.442
Glu F3C	$O_C - N$	0.103	0.127	0.242	0.439
Glu SPC/E	$O_C - N$	0.115	0.144	0.284	0.505
Glu TIP3P	$O_C - N$	0.107	0.134	0.254	0.449
Glu F3C	$O_{CE} - Na^+$	0.096	0.112	0.201	0.375
Glu SPC/E	$O_{CE} - Na^+$	0.102	0.119	0.219	0.420
Glu TIP3P	$O_{CE} - Na^+$	0.100	0.119	0.213	0.389
Glu F3C	$O_C - Na^+$	0.097	0.113	0.203	0.383
Glu SPC/E	$O_C - Na^+$	0.098	0.114	0.208	0.395
Glu TIP3P	$O_C - Na^+$	0.101	0.118	0.207	0.371
Glu F3C	$O_{CE} - O_W$	0.083	0.086	0.114	0.180
Glu SPC/E	$O_{CE} - O_W$	0.084	0.087	0.119	0.196
Glu TIP3P	$O_{CE} - O_W$	0.083	0.086	0.111	0.169
Glu F3C	$O_C - O_W$	0.083	0.086	0.114	0.180
Glu SPC/E	$O_C - O_W$	0.083	0.087	0.119	0.196
Glu TIP3P	$O_C - O_W$	0.083	0.086	0.111	0.168
Glu F3C	$N - O_W$	0.080	0.082	0.093	0.113
Glu SPC/E	$N - O_W$	0.080	0.082	0.094	0.116
Glu TIP3P	$N - O_W$	0.080	0.081	0.090	0.105
Glu F3C	$Na^+ - O_W$	0.080	0.082	0.093	0.113
Glu SPC/E	$Na^+ - O_W$	0.080	0.082	0.094	0.116
Glu TIP3P	$Na^+ - O_W$	0.080	0.081	0.090	0.105
Glu F3C	$O_W - O_W$	0.080	0.082	0.093	0.113
Glu SPC/E	$O_W - O_W$	0.080	0.082	0.094	0.116
Glu TIP3P	$O_W - O_W$	0.080	0.081	0.090	0.105
Pure SPC/E	$O_W - O_W$	0.082	0.085	0.104	0.143

Table B.2: Bond lifetimes for each of the classical glutamate systems and the pure SPC/E water system.

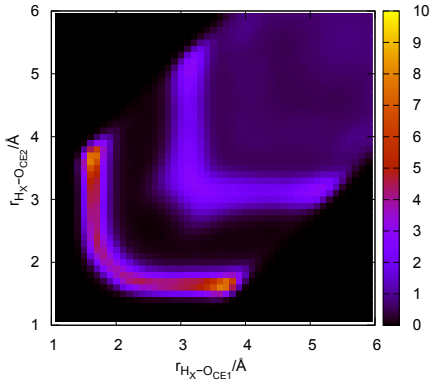
Temperer index	Temperature (K)
1	300
2	302
3	308
4	314
5	318
6	323
7	329
8	335
9	342
10	348
11	355
12	360
13	367
14	373
15	380
16	386
17	397
18	403
19	412
20	419
21	426
22	433
23	439
24	446
25	450
26	461
27	467
28	473
29	480
30	488
31	493
32	500

Table B.3: Temperatures of each of the 32 temperers in the Parallel Tempering simulation.

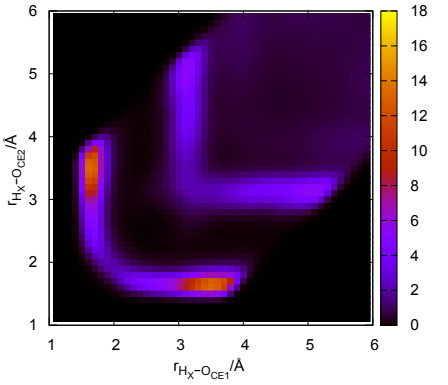


(a) Original system $H_X - O_{CE}$

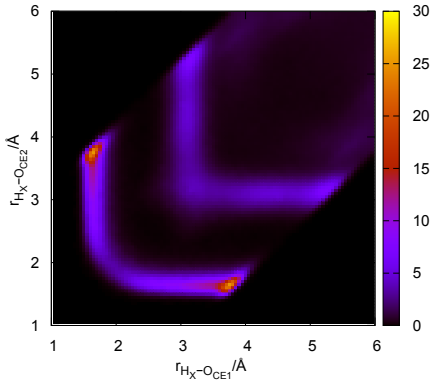
(b) Initial Configuration $H_X - O_{CE}$



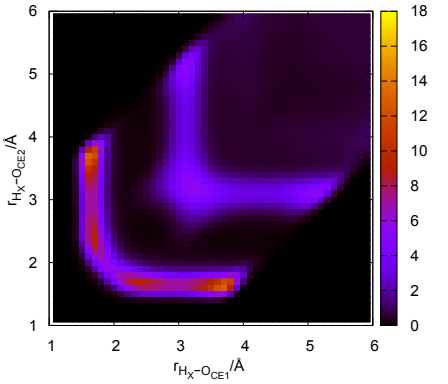
(c) Temp12 $H_X - O_{CE}$



(d) Temp21 $H_X - O_{CE}$

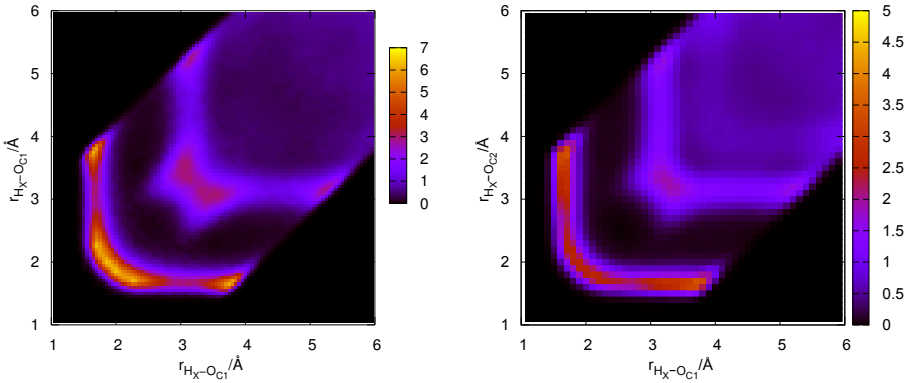


(e) Temp22 $H_X - O_{CE}$

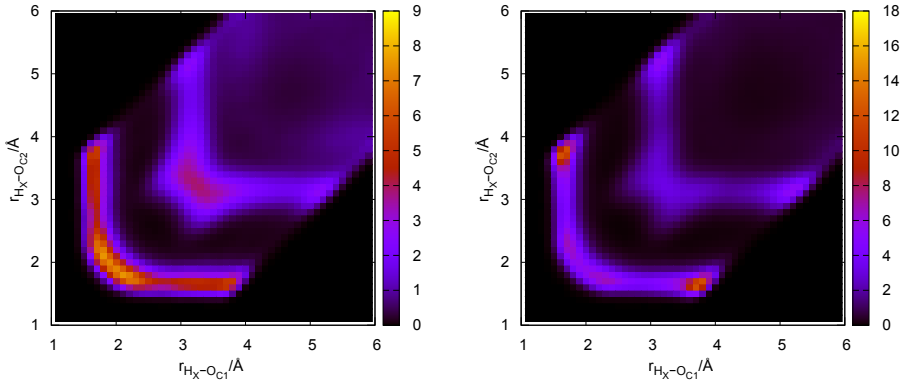


(f) Temp32 $H_X - O_{CE}$

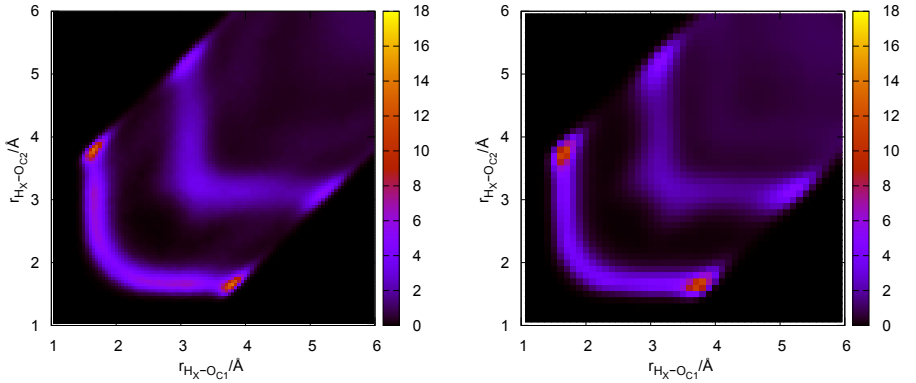
Figure B.1: 2D radial-radial distribution functions of the Glu-Glu interactions between H_X and O_{CE} atoms.



(a) Original system $H_X - O_C$ (b) Initial Configuration $H_X - O_C$

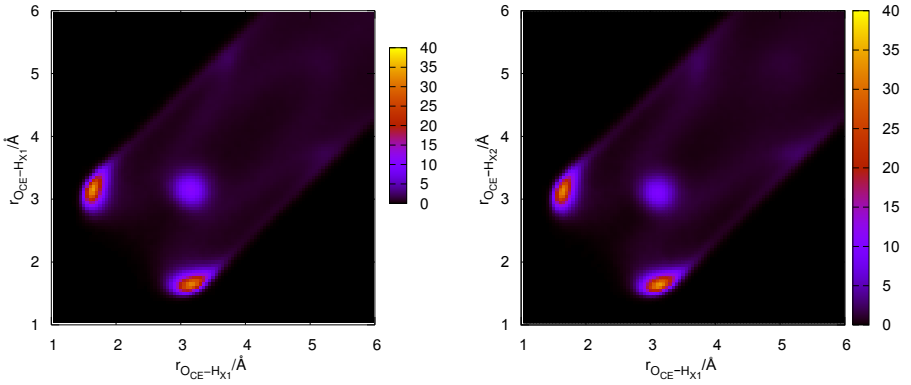


(c) Temp12 $H_X - O_C$ (d) Temp21 $H_X - O_C$

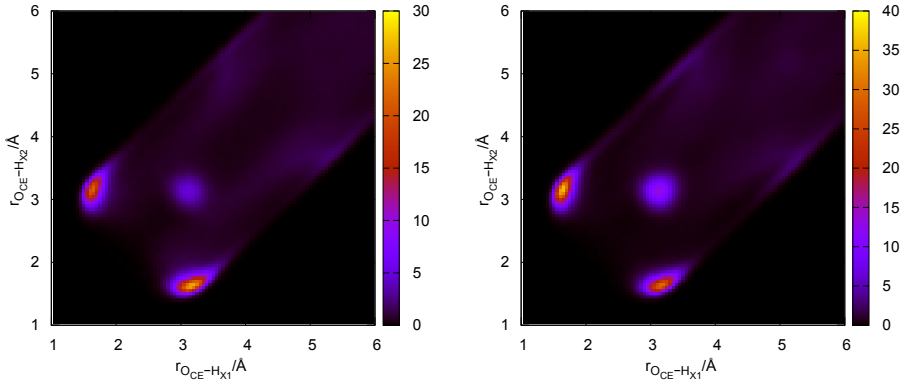


(e) Temp22 $H_X - O_C$ (f) Temp32 $H_X - O_C$

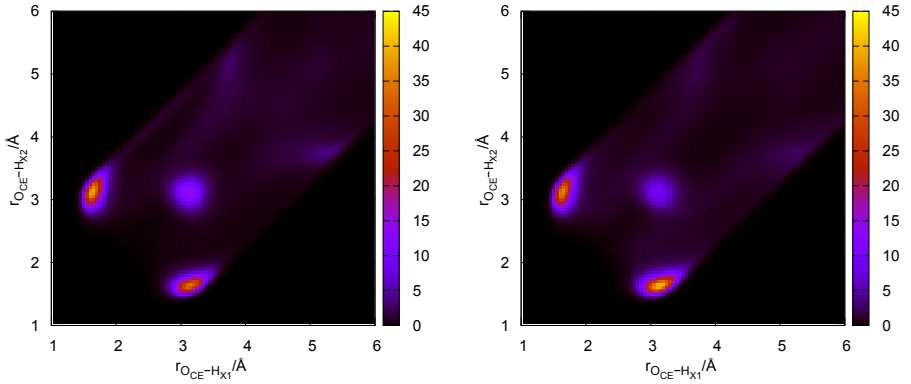
Figure B.2: 2D radial-radial distribution functions of the Glu-Glu interactions between H_X and O_C atoms.



(a) Original system $O_{CE} - H_X$ (b) Initial Configuration $O_{CE} - H_X$

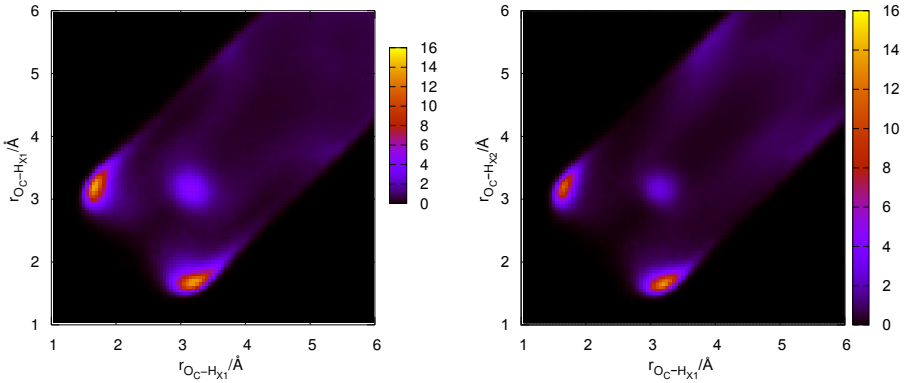


(c) Temp12 $O_{CE} - H_X$ (d) Temp21 $O_{CE} - H_X$

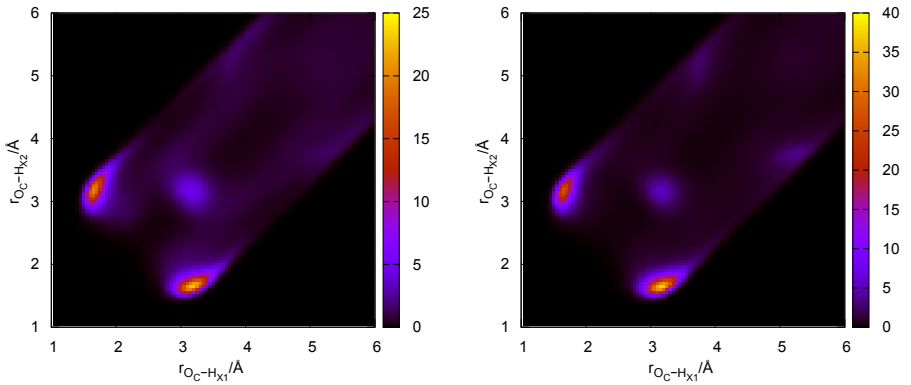


(e) Temp22 $O_{CE} - H_X$ (f) Temp32 $O_{CE} - H_X$

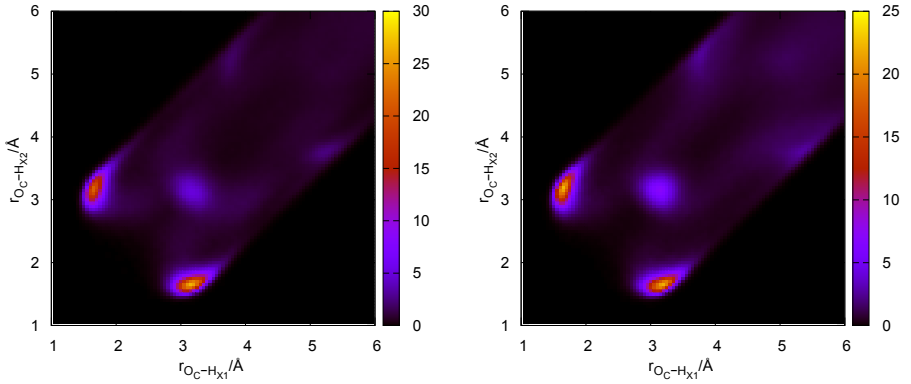
Figure B.3: 2D radial-radial distribution functions of the Glu-Glu interactions between O_{CE} and H_X atoms.



(a) Original system $O_C - H_X$ (b) Initial Configuration $O_C - H_X$

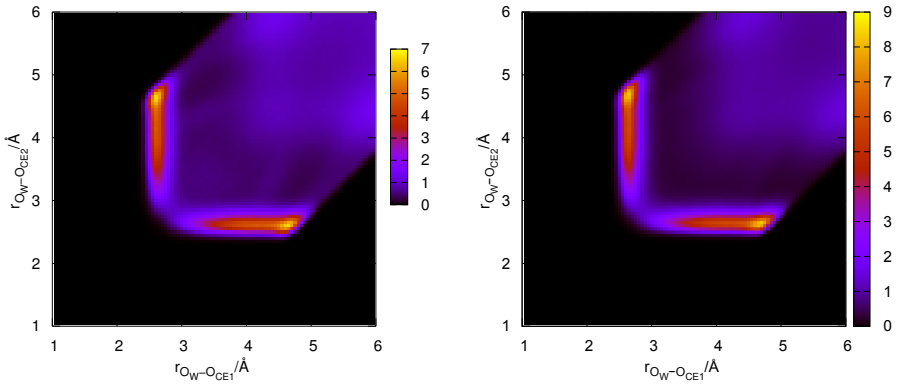


(c) Temp12 $O_C - H_X$ (d) Temp21 $O_C - H_X$

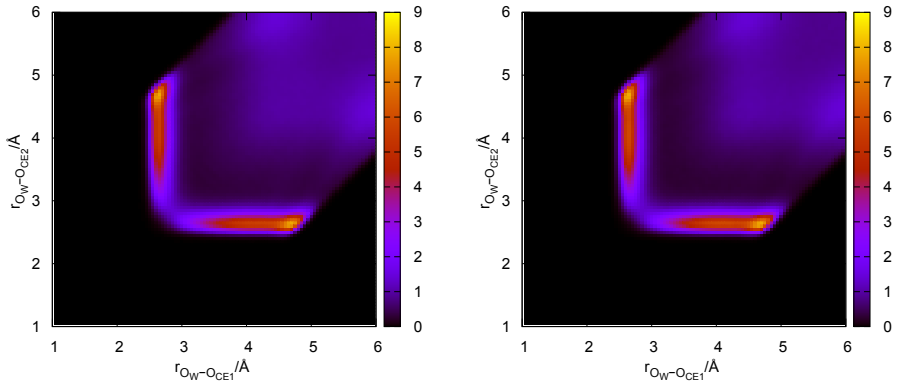


(e) Temp22 $O_C - H_X$ (f) Temp32 $O_C - H_X$

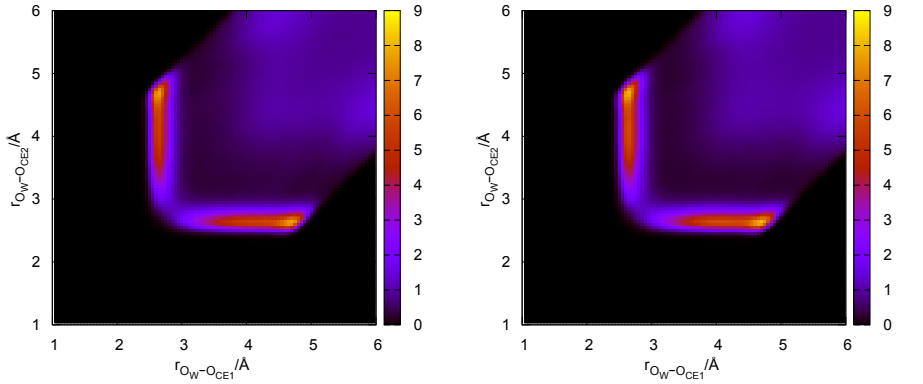
Figure B.4: 2D radial-radial distribution functions of the Glu-Glu interactions between O_C and H_X atoms.



(a) Original system $O_W - O_{CE}$ (b) Initial Configuration $O_W - O_{CE}$

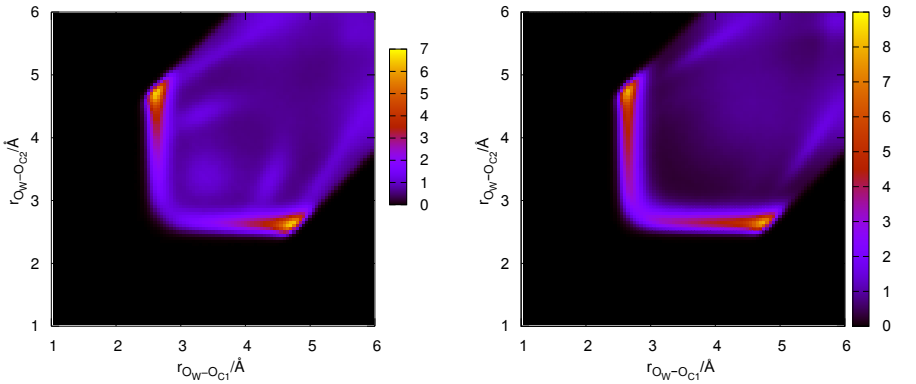


(c) Temp12 $O_W - O_{CE}$ (d) Temp21 $O_W - O_{CE}$

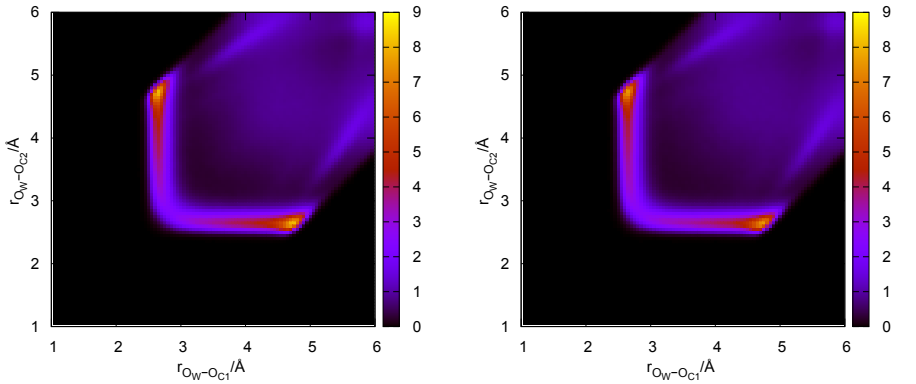


(e) Temp22 $O_W - O_{CE}$ (f) Temp32 $O_W - O_{CE}$

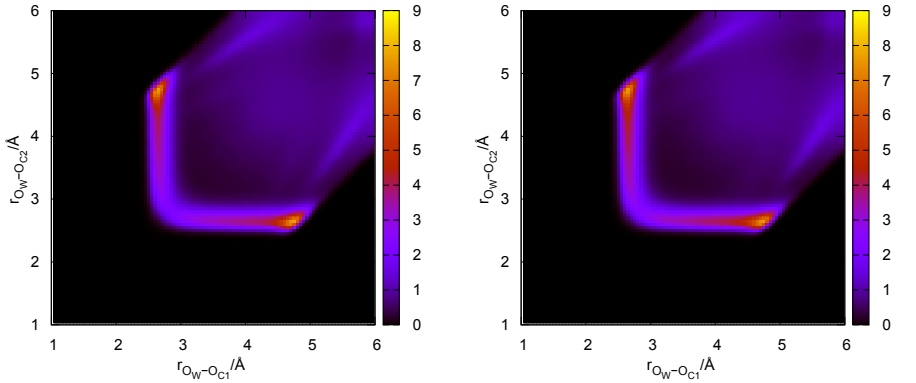
Figure B.5: 2D radial-radial distribution functions of the glu-water interactions between O_W (water) and O_{CE} (glutamate) atoms for the original system from chapter 3, and for the temperer 22 system.



(a) Original system $O_W - O_C$ (b) Initial Configuration $O_W - O_C$

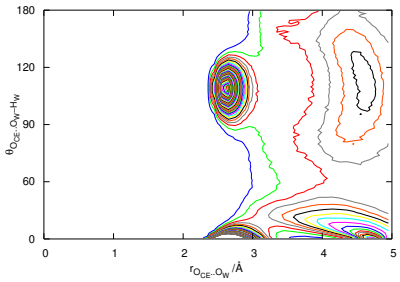


(c) Temp12 $O_W - O_C$ (d) Temp21 $O_W - O_C$

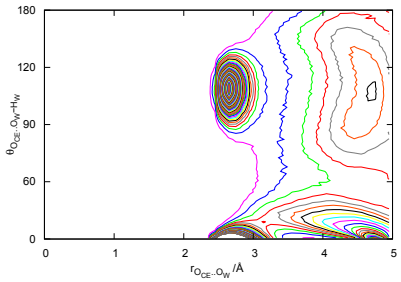


(e) Temp22 $O_W - O_C$ (f) Temp32 $O_W - O_C$

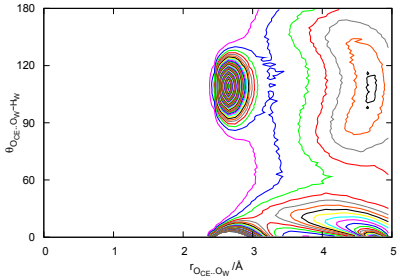
Figure B.6: 2D radial-radial distribution functions of the glu-water interactions between O_W (water) and O_C (glutamate) atoms for the original system from chapter 3, and for the temperer 22 system.



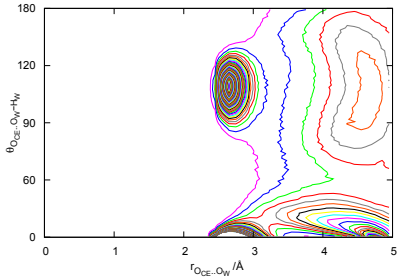
(a) Original system $O_{CE}\cdot\cdot O_W - H_W$



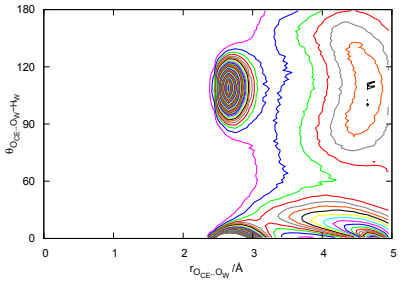
(b) Initial Configuration $O_{CE}\cdot\cdot O_W - H_W$



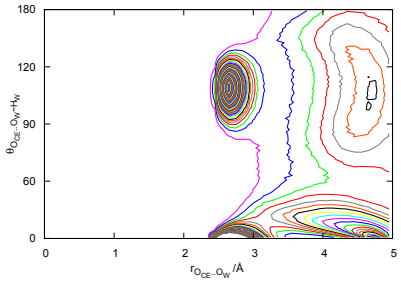
(c) Temperer 12 $O_{CE}\cdot\cdot O_W - H_W$



(d) Temperer 21 $O_{CE}\cdot\cdot O_W - H_W$

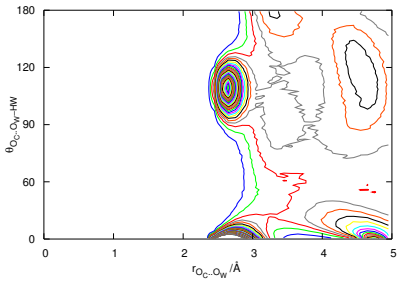


(e) Temperer 22 $O_{CE}\cdot\cdot O_W - H_W$

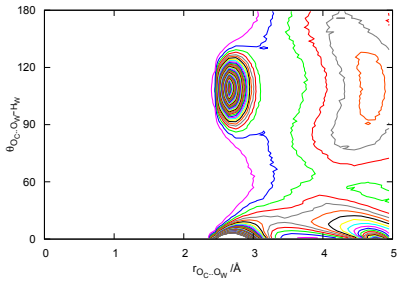


(f) Temperer 32 $O_{CE}\cdot\cdot O_W - H_W$

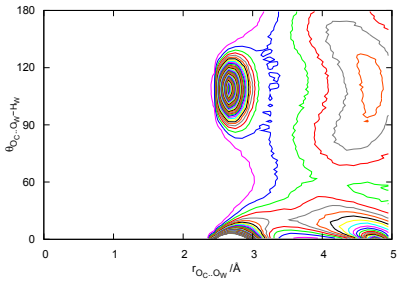
Figure B.7: 2D radial-angular distribution functions of the glu-water interactions between O_{CE} atom and the water molecule, correlating the $O_{CE} - O_W$ radial distance with the $O_{CE} - O_W - H_W$ angle from the original simulation and the temperer 22 system.



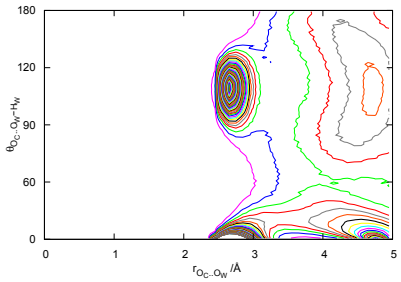
(a) Original system $O_C \cdots O_W - H_W$



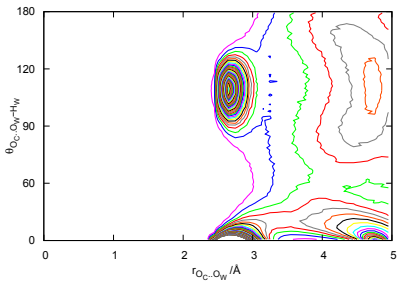
(b) Initial Configuration $O_C \cdots O_W - H_W$



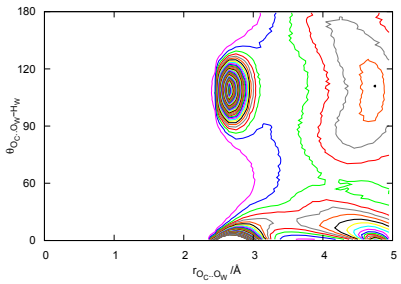
(c) Temperer 12 $O_C \cdots O_W - H_W$



(d) Temperer 21 $O_C \cdots O_W - H_W$

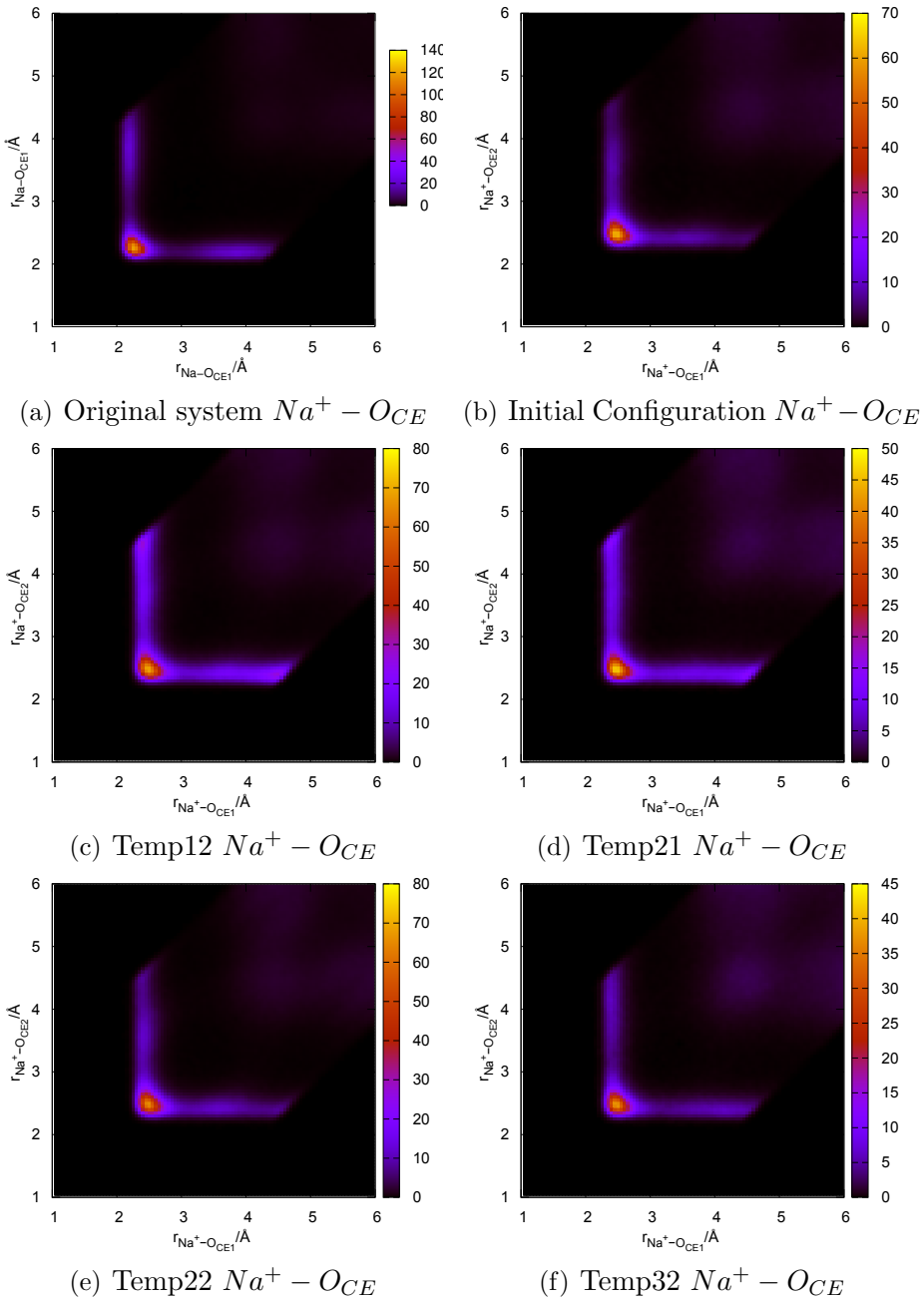


(e) Temperer 22 $O_C \cdots O_W - H_W$



(f) Temperer 32 $O_C \cdots O_W - H_W$

Figure B.8: 2D radial-angular distribution functions of the glu-water interactions between OC atom and the water molecule, correlating the $O_C - O_W$ radial distance with the $O_C - O_W - H_W$ angle from the original simulation and the temperer 22 system.

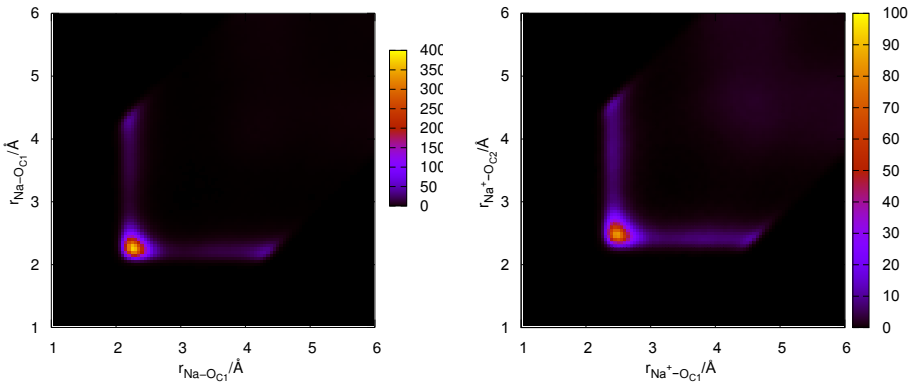


(a) Original system $Na^+ - O_{CE}$ (b) Initial Configuration $Na^+ - O_{CE}$

(c) Temp12 $Na^+ - O_{CE}$ (d) Temp21 $Na^+ - O_{CE}$

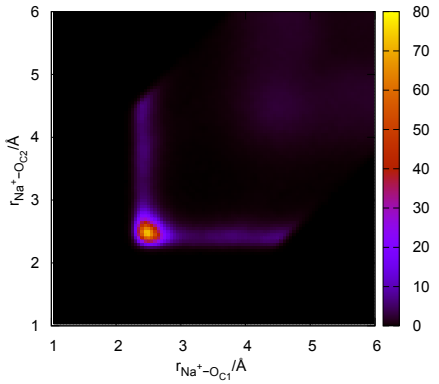
(e) Temp22 $Na^+ - O_{CE}$ (f) Temp32 $Na^+ - O_{CE}$

Figure B.9: 2D radial-radial distribution functions of the sodium-glu interactions between Na^+ (sodium ion) and O_{CE} (glutamate) atoms.

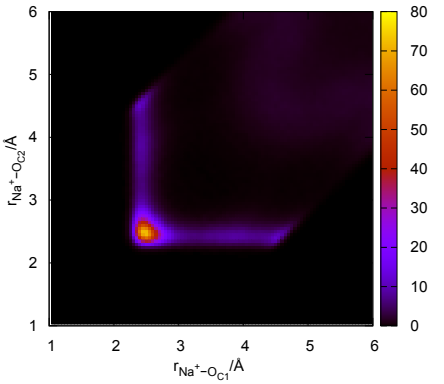


(a) Original system $Na^+ - O_C$

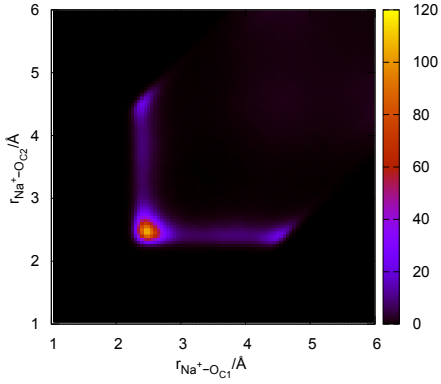
(b) Initial Configuration $Na^+ - O_C$



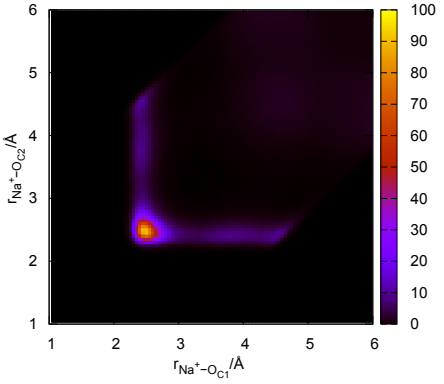
(c) Temp12 $Na^+ - O_C$



(d) Temp21 $Na^+ - O_C$



(e) Temp22 $Na^+ - O_C$



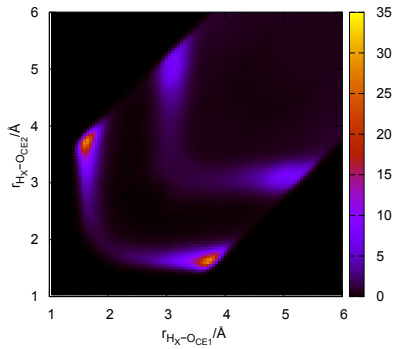
(f) Temp32 $Na^+ - O_C$

Figure B.10: 2D radial-radial distribution functions of the sodium-glu interactions between Na^+ (sodium ion) and O_C) atoms.

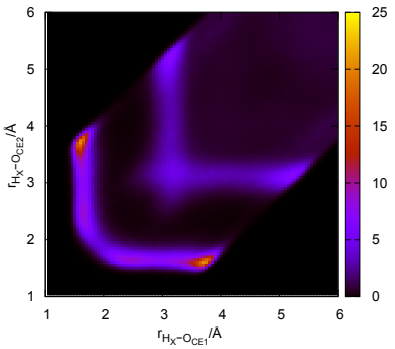
Appendix C

Additional information on the structure of GPE

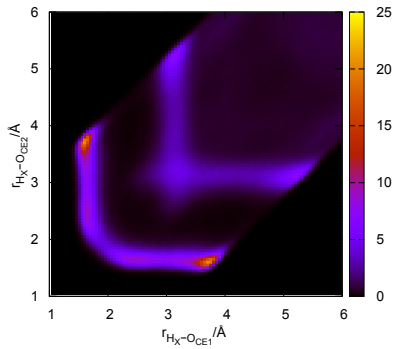
The following material is supplementary to discussion in chapter 5.



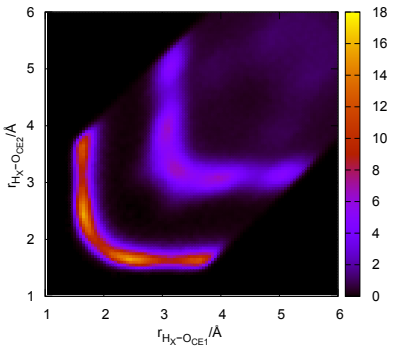
(a) 1:29 A Closed



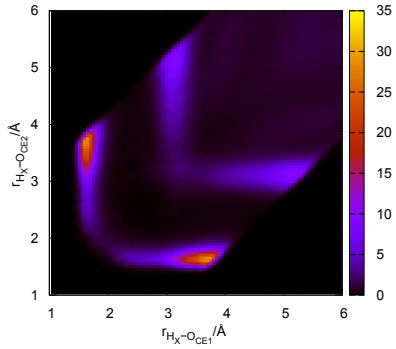
(b) 1:29 A Open



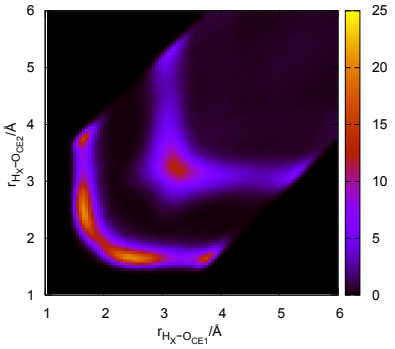
(c) 1:29 D Open



(d) 1:61 A Closed

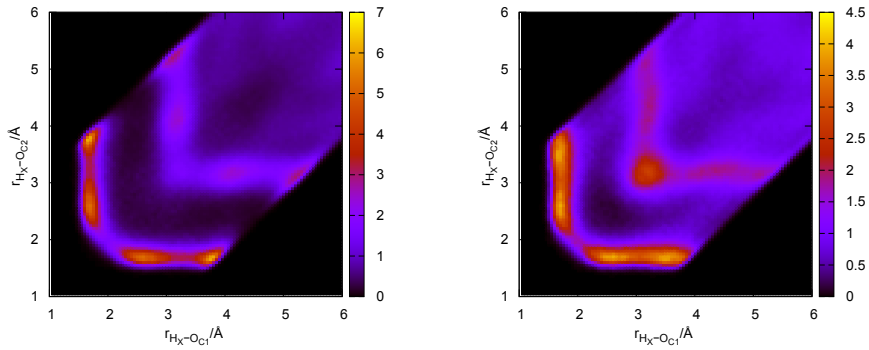


(e) 1:61 A Open



(f) 1:61 D Open

Figure C.1: 2D radial-radial distribution functions for the GPE-GPE interactions $H_X - O_{CE}$ in 1:1:29 and 1:1:61 solutions of GPE.



(a) 1:29 A Closed

(b) 1:29 A Open

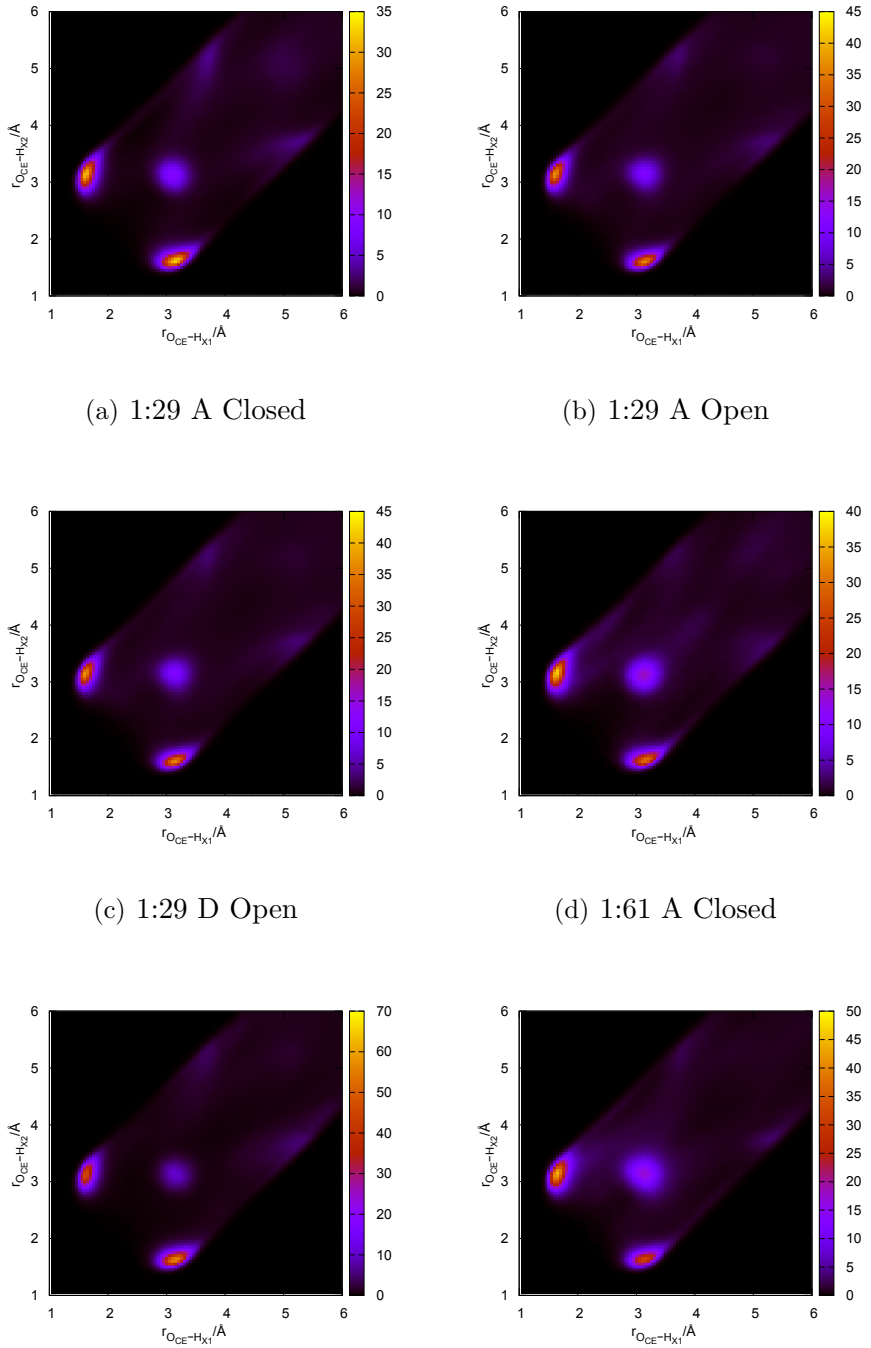
(c) 1:29 D Open

(d) 1:61 A Closed

(e) 1:61 A Open

(f) 1:61 D Open

Figure C.2: 2D radial-radial distribution functions for the GPE-GPE interactions $H_X - O_C$ in 1:1:29 and 1:1:61 solutions of GPE.



(a) 1:29 A Closed

(b) 1:29 A Open

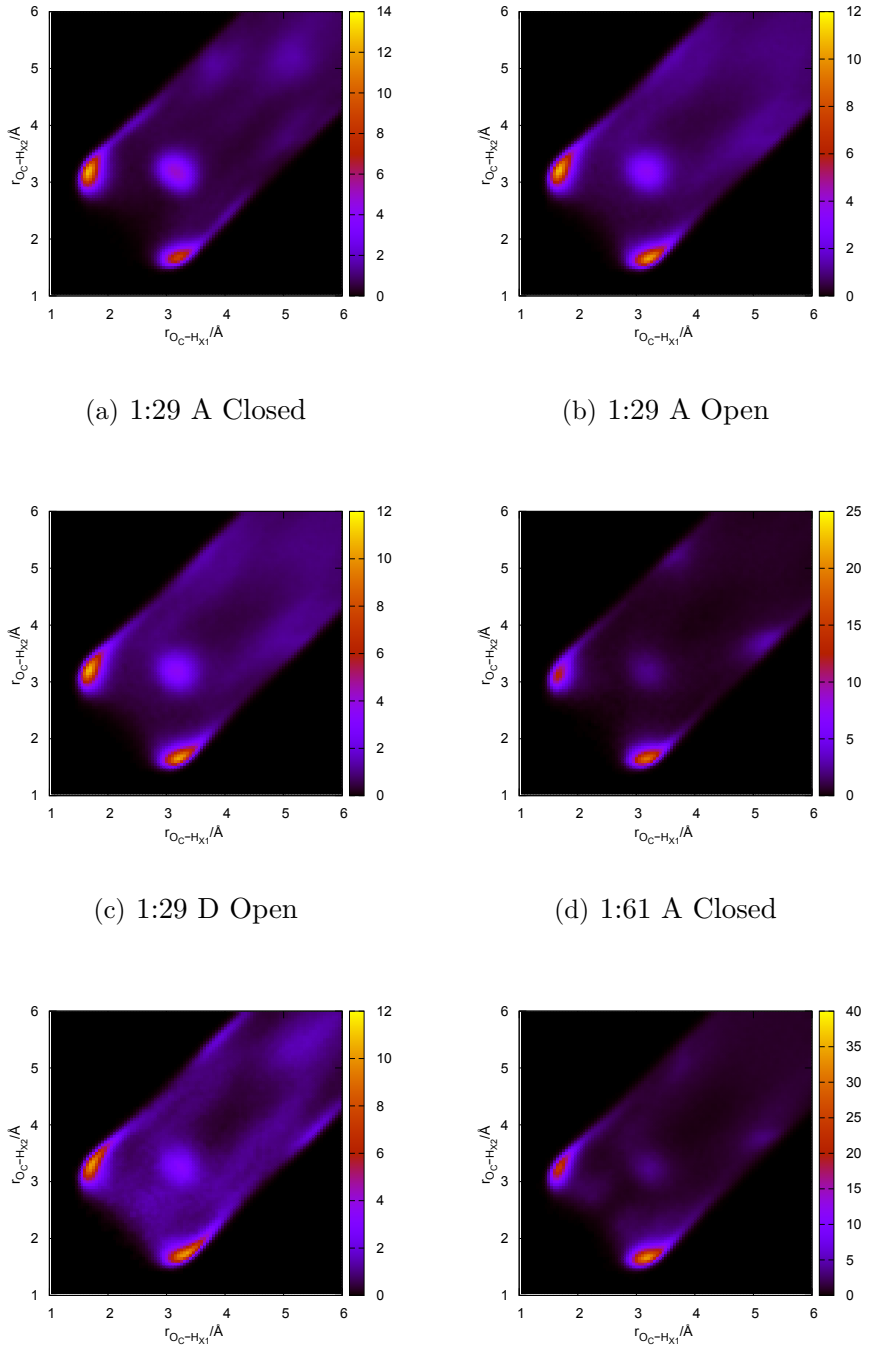
(c) 1:29 D Open

(d) 1:61 A Closed

(e) 1:61 A Open

(f) 1:61 D Open

Figure C.3: 2D radial-radial distribution functions for the GPE-GPE interactions $O_{CE} - H_X$ in 1:1:29 and 1:1:61 solutions of GPE. These show no bifurcated bonds forming on a single O_{CE} atom.



(a) 1:29 A Closed

(b) 1:29 A Open

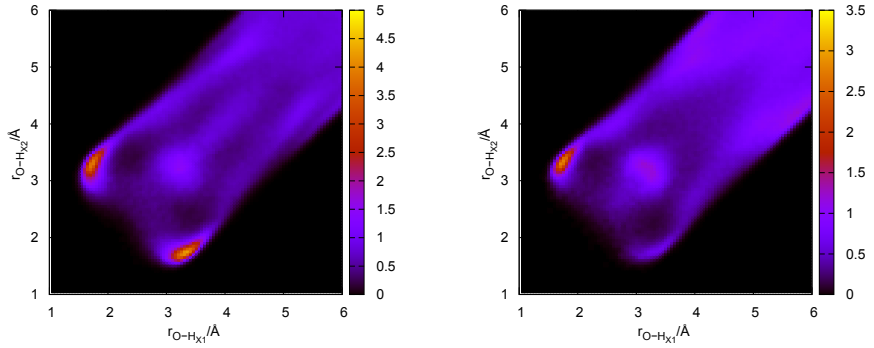
(c) 1:29 D Open

(d) 1:61 A Closed

(e) 1:61 A Open

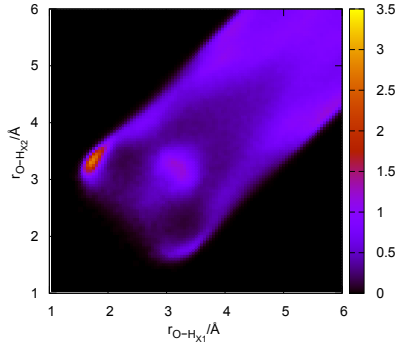
(f) 1:61 D Open

Figure C.4: 2D radial-radial distribution functions for the GPE-GPE interactions $O_C - H_X$ in 1:1:29 and 1:1:61 solutions of GPE. These show no bifurcated bonds forming on a single O_C atom.

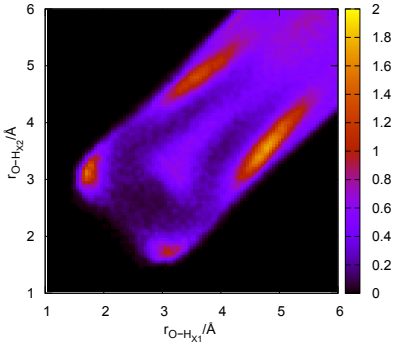


(a) 1:29 A Closed

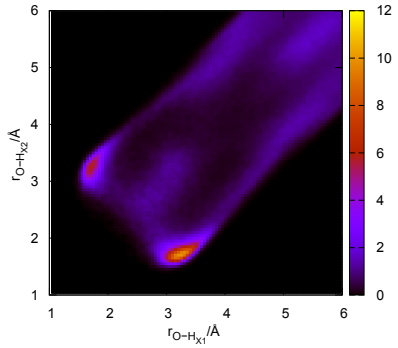
(b) 1:29 A Open



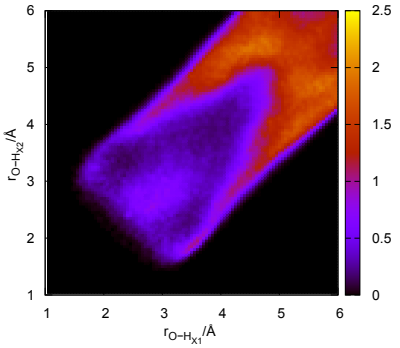
(c) 1:29 D Open



(d) 1:61 A Closed



(e) 1:61 A Open

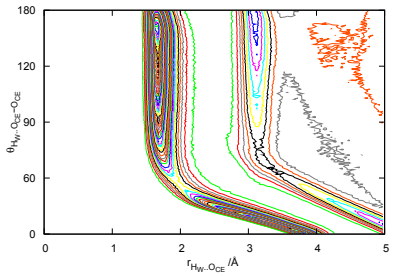


(f) 1:61 D Open

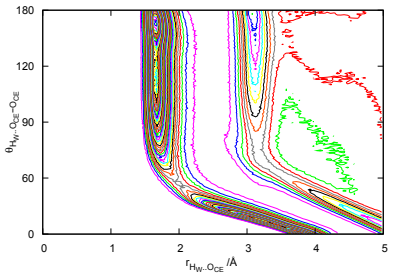
Figure C.5: 2D radial-radial distribution functions for the GPE-GPE interactions $O-H_X$ in 1:1:29 and 1:1:61 solutions of GPE. These show no bifurcated bonds forming where one O atom bonds to two H_X atoms.

Interaction	r_{min} (Å)	r_{cut} (Å)	$g(r_{min})$	n^{α}_{β}	n^{β}_{α}
1:1:29 Closed A					
$H_X - O_{CE}$	2.415	2.3	0.193678	0.255967647	0.383951471
$H_X - O_C$	2.295	2.3	0.429364	0.15655	0.234825
$H_X - O$	2.265	2.3	0.25635	0.060383333	0.090575
$N - O$CE	3.345	3.6	0.339139	0.860120588	0.430060294
$N - O$C	3.765	3.6	0.632912	0.674582353	0.337291176
$N - O$	3.705	3.6	0.479875	0.363270588	0.181635294
1:1:29 Open A					
$H_X - O_{CE}$	2.475	2.3	0.310346	0.250193137	0.375289706
$H_X - O_C$	2.235	2.3	0.736153	0.226941176	0.340411765
$H_X - O$	2.265	2.3	0.461425	0.123817647	0.185726471
$N - O$CE	3.495	3.6	0.339078	0.86045	0.430225
$N - O$C	3.705	3.6	0.797041	1.00105	0.500525
$N - O$	3.555	3.6	0.471709	0.603847059	0.301923529
1:1:29 Open D					
$H_X - O_{CE}$	2.355	2.3	0.538796	0.291876471	0.437814706
$H_X - O_C$	2.415	2.3	0.52701	0.148741176	0.223111765
$H_X - O$	2.145	2.3	0.189098	0.03732549	0.055988235
$N - O$CE	3.585	3.6	0.618789	1.086147059	0.543073529
$N - O$C	3.705	3.6	0.538297	0.66315	0.331575
$N - O$	3.495	3.6	0.317876	0.240947059	0.120473529
1:1:61 Closed A					
$H_X - O_{CE}$	2.475	2.3	0.73841	0.106020588	0.159030882
$H_X - O_C$	2.445	2.3	0.398881	0.170070588	0.255105882
$H_X - O$	2.355	2.3	0.111731	0.047926471	0.071889706
$N - O$CE	3.855	3.6	0.637597	0.401629412	0.200814706
$N - O$C	3.555	3.6	0.41832	0.192938235	0.096469118
$N - O$	3.165	3.6	0.310081	0.053082353	0.026541176
1:1:61 Open A					
$H_X - O_{CE}$	2.325	2.3	0.310659	0.104309804	0.156464706
$H_X - O_C$	2.295	2.3	0.73062	0.04795	0.071925
$H_X - O$	2.295	2.3	0.35481	0.027031373	0.040547059
$N - O$CE	3.525	3.6	0.462005	0.352664706	0.176332353
$N - O$C	3.975	3.6	0.603204	0.239941176	0.119970588
$N - O$	3.645	3.6	0.511133	0.138244118	0.069122059
1:1:61 Open D					
$H_X - O_{CE}$	2.415	2.3	1.17009	0.133603922	0.200405882
$H_X - O_C$	2.445	2.3	1.04061	0.089858824	0.134788235
$H_X - O$	1.935	2.3	0.144563	0.005112745	0.007669118
$N - O$CE	3.945	3.6	0.689933	0.530708824	0.265354412
$N - O$C	3.915	3.6	0.795339	0.393570588	0.196785294
$N - O$	3.465	3.6	0.511874	0.067758824	0.033879412

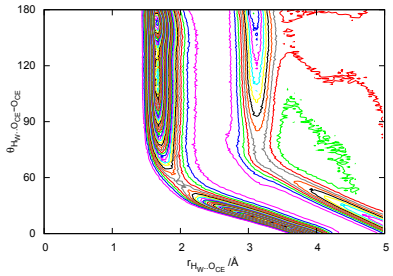
Table C.1: Co-ordination numbers, n^{α}_{β}, for the GPE-GPE bond sites.



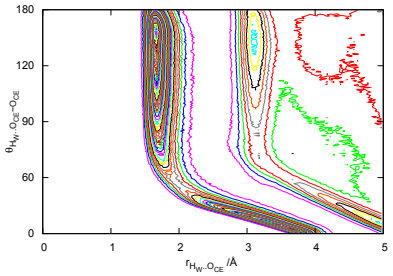
(a) 1:1:29 Closed A



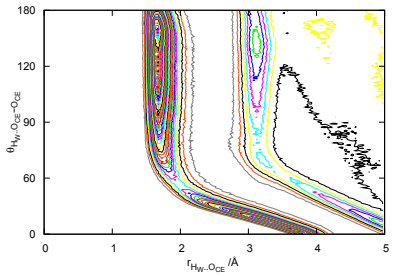
(b) 1:1:29 Open A



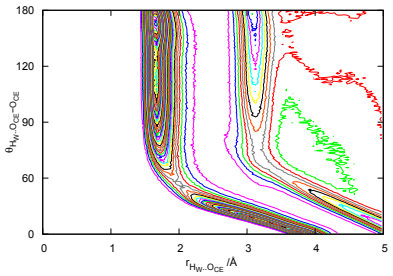
(c) 1:1:29 Open D



(d) 1:1:61 Closed A

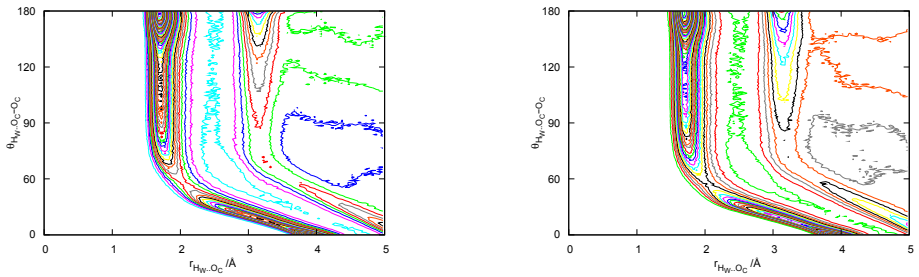


(e) 1:1:61 Open A



(f) 1:1:61 Open D

Figure C.6: 2D radial-angular distribution functions for the GPE-GPE interactions between water and the side-chain carboxyl oxygens (O_{CE}). This correlates the $r_{H_W-O_{CE1}}$ distance with the $H_W - \widehat{O_{CE1}} - O_{CE2}$ angle.



(a) 1:1:29 Closed A

(b) 1:1:29 Open A

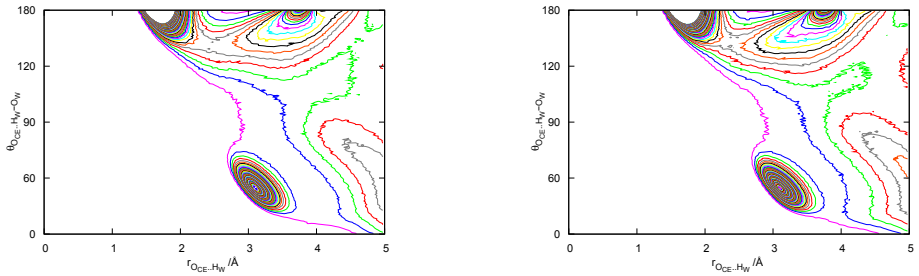
(c) 1:1:29 Open D

(d) 1:1:61 Closed A

(e) 1:1:61 Open A

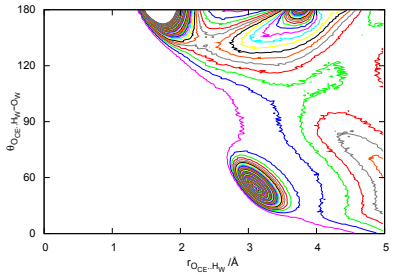
(f) 1:1:61 Open D

Figure C.7: 2D radial-angular distribution functions for the water-GPE interactions between water and the side-chain carboxyl oxygens (O_C). This correlates the $r_{H_W-O_{C1}}$ distance with the $H_W - \widehat{O_{C1} - O_{C2}}$ angle.

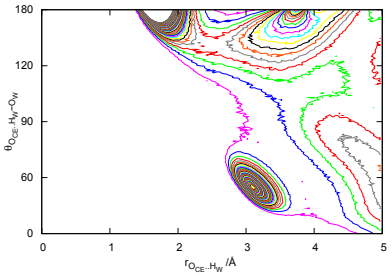


(a) 1:1:29 Closed A

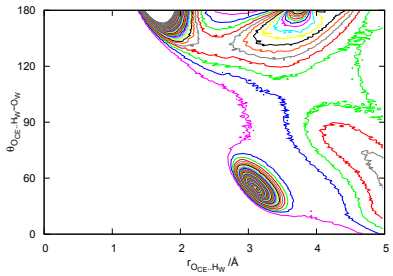
(b) 1:1:29 Open A



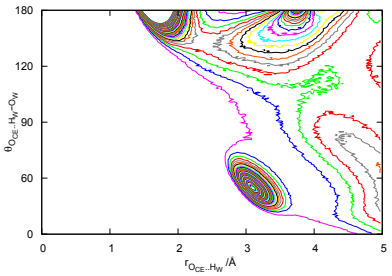
(c) 1:1:29 Open D



(d) 1:1:61 Closed A

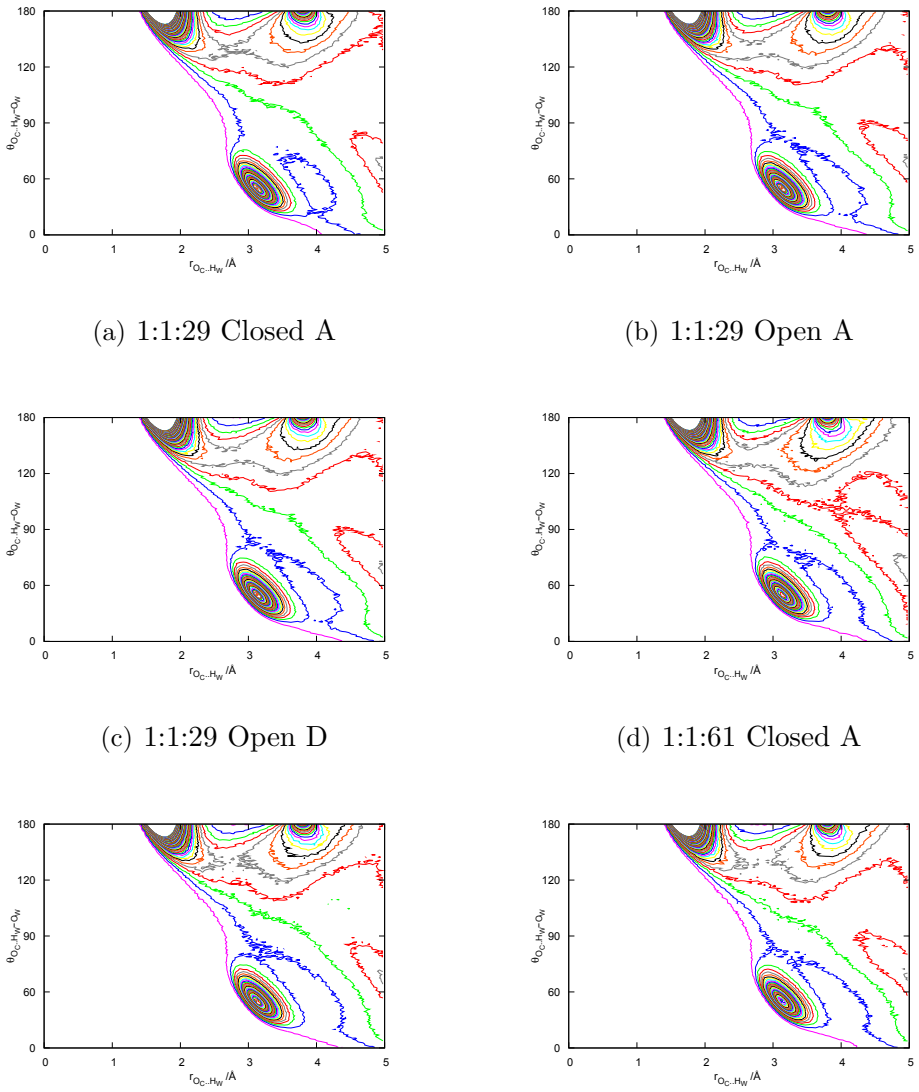


(e) 1:1:61 Open A



(f) 1:1:61 Open D

Figure C.8: 2D radial-angular distribution functions for the water-GPE interactions between water and the side-chain carboxyl oxygens (O_{CE}). This correlates the $r_{O_{CE}-H_W}$ distance with the $O_{CE} - \widehat{H_W} - O_W$ angle.



(a) 1:1:29 Closed A

(b) 1:1:29 Open A

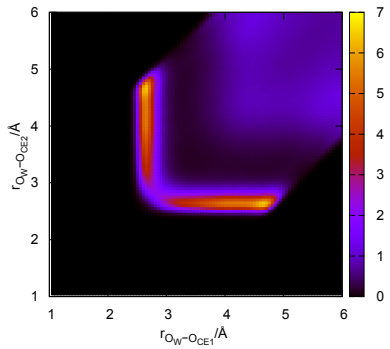
(c) 1:1:29 Open D

(d) 1:1:61 Closed A

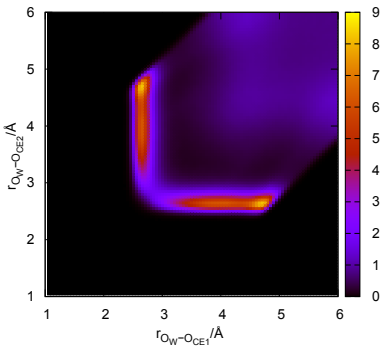
(e) 1:1:61 Open A

(f) 1:1:61 Open D

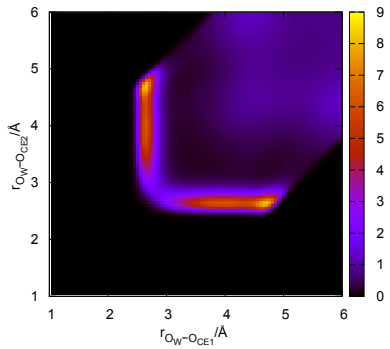
Figure C.9: 2D radial-angular distribution functions for the GPE-water interactions between water and the side-chain carboxyl oxygens (O_C). This correlates the $r_{O_C-H_W}$ distance with the $O_C - \widehat{H_W - O_W}$ angle.



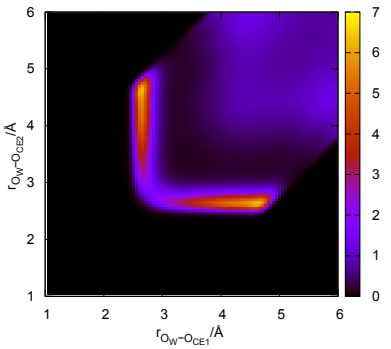
(a) 1:29 A Closed



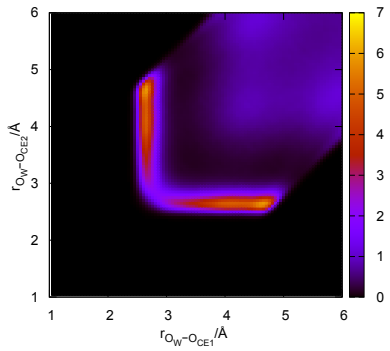
(b) 1:29 A Open



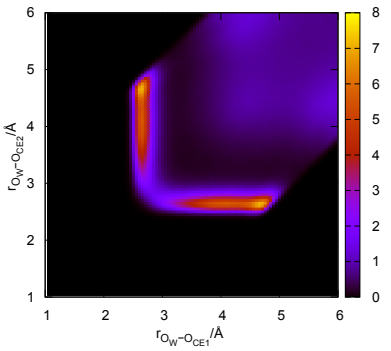
(c) 1:29 D Open



(d) 1:61 A Closed

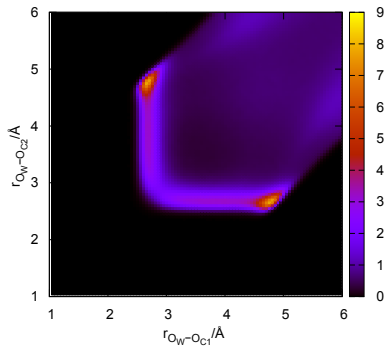


(e) 1:61 A Open

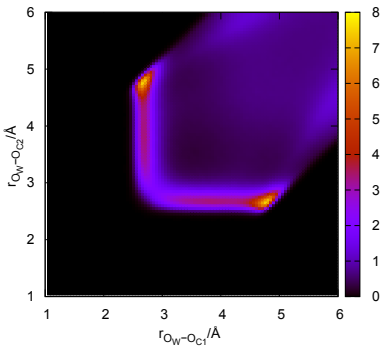


(f) 1:61 D Open

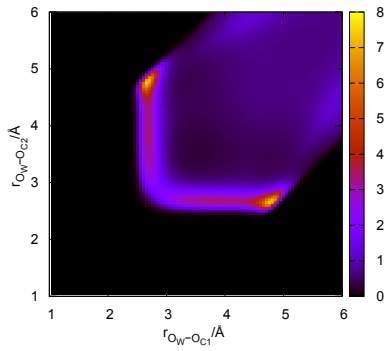
Figure C.10: 2D radial-radial distribution functions for the water-GPE interactions $O_W - O_{CE}$ in 1:1:29 and 1:1:61 solutions of GPE.



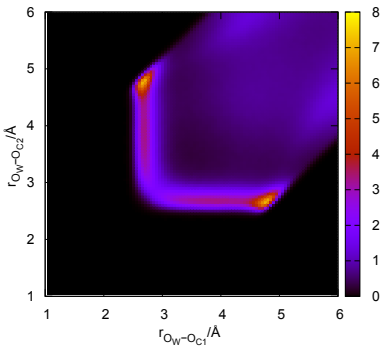
(a) 1:29 A Closed



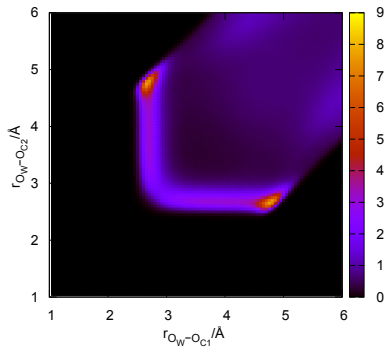
(b) 1:29 A Open



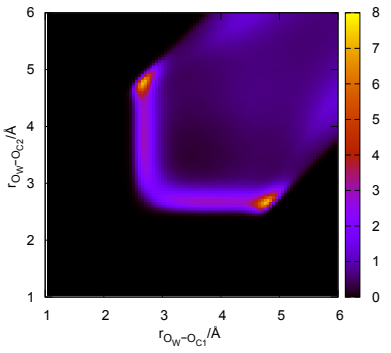
(c) 1:29 D Open



(d) 1:61 A Closed

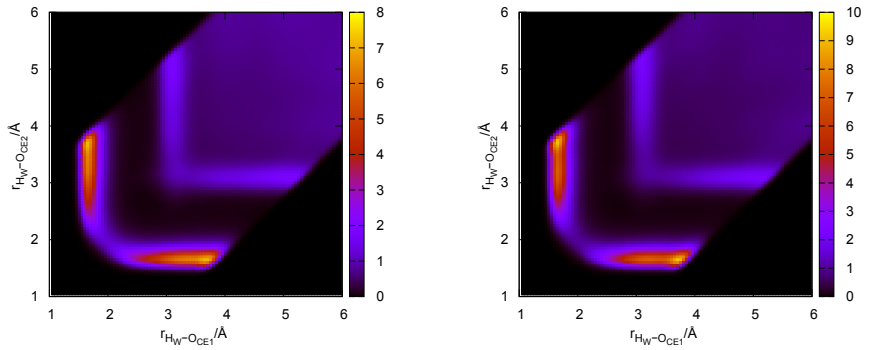


(e) 1:61 A Open



(f) 1:61 D Closed

Figure C.11: 2D radial-radial distribution functions for the water-GPE interactions $O_W - O_C$ in 1:1:29 and 1:1:61 solutions of GPE.



(a) 1:29 A Closed

(b) 1:29 A Open

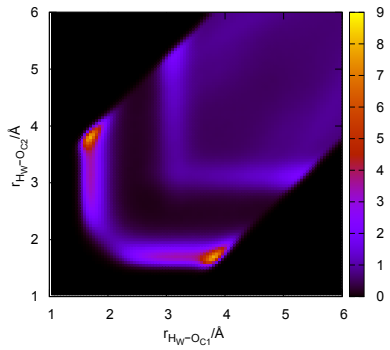
(c) 1:29 D Open

(d) 1:61 A Closed

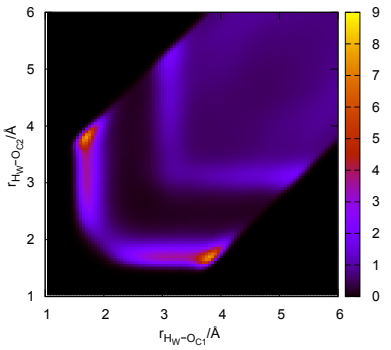
(e) 1:61 A Open

(f) 1:61 D Open

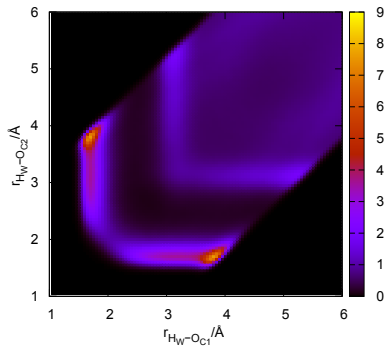
Figure C.12: 2D radial-radial distribution functions for the water-GPE interactions $H_W - O_{CE}$ in 1:1:29 and 1:1:61 solutions of GPE.



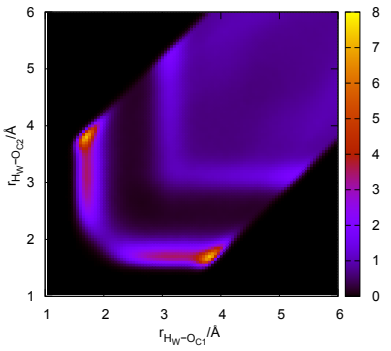
(a) 1:29 A Closed



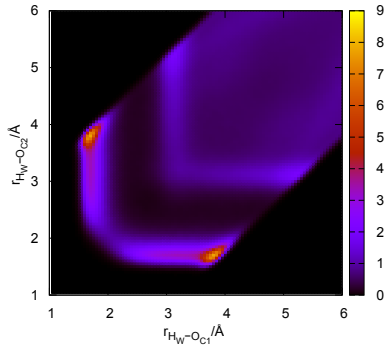
(b) 1:29 A Open



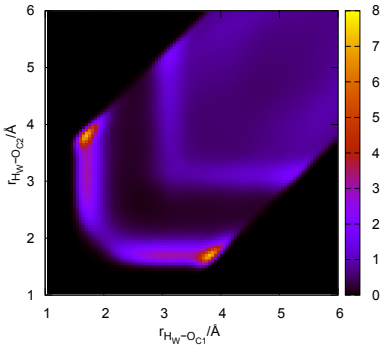
(c) 1:29 D Open



(d) 1:61 A Closed

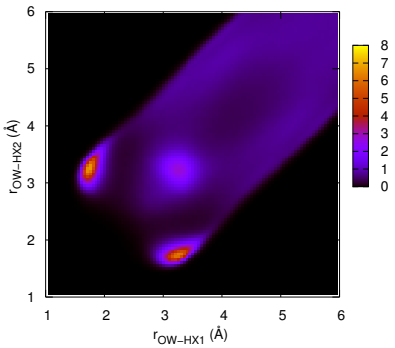


(e) 1:61 A Open

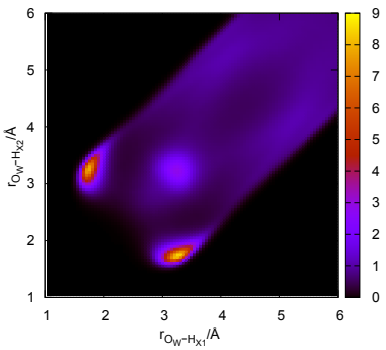


(f) 1:61 D Closed

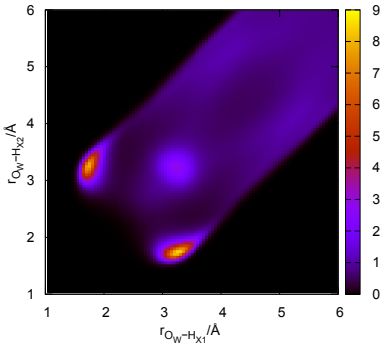
Figure C.13: 2D radial-radial distribution functions for the water-GPE interactions $H_W - O_C$ in 1:1:29 and 1:1:61 solutions of GPE.



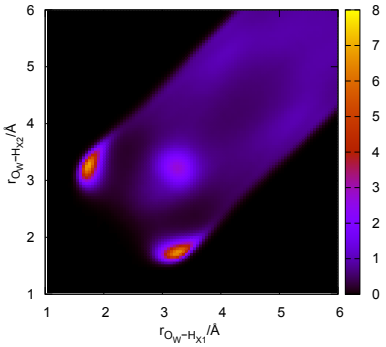
(a) 1:29 A Closed



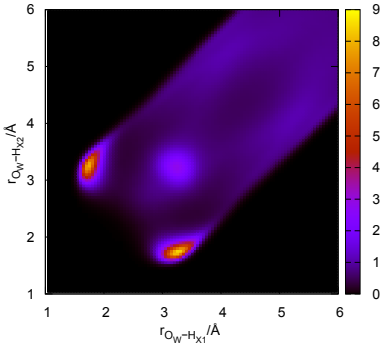
(b) 1:29 A Open



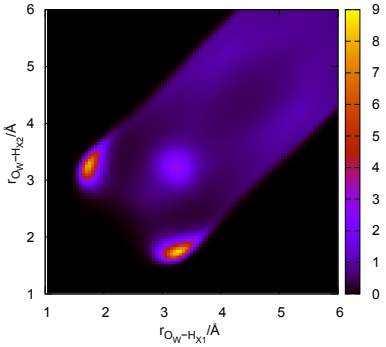
(c) 1:29 D Open



(d) 1:61 A Closed

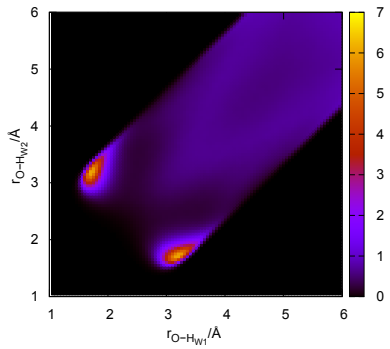


(e) 1:61 A Open

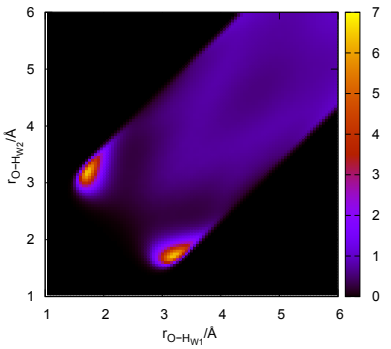


(f) 1:61 D Open

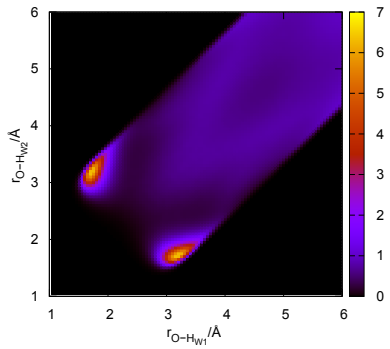
Figure C.14: 2D radial-radial distribution functions for the water-GPE interactions $O_W - H_X$ in 1:1:29 and 1:1:61 solutions of GPE.



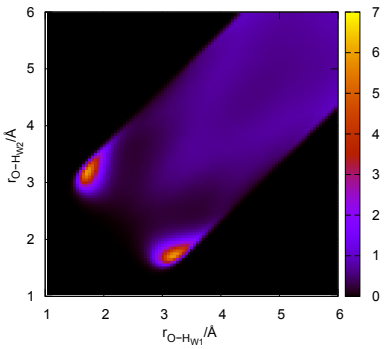
(a) 1:29 A Closed



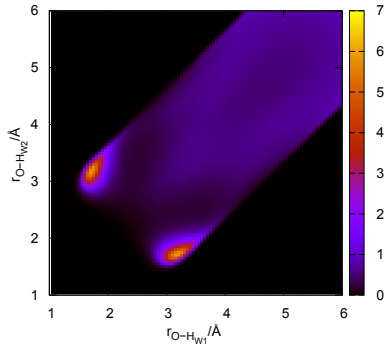
(b) 1:29 A Open



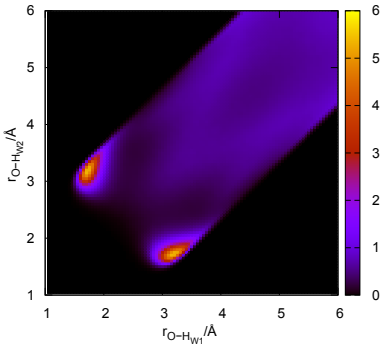
(c) 1:29 D Open



(d) 1:61 A Closed



(e) 1:61 A Open

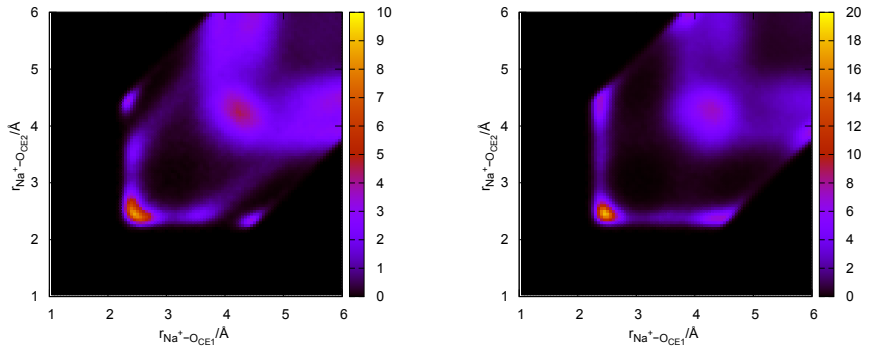


(f) 1:61 D Open

Figure C.15: 2D radial-radial distribution functions for the GPE-water interactions $O - H_W$ in 1:1:29 and 1:1:61 solutions of GPE.

Interaction	r_{min} (Å)	r_{cut} (Å)	$g(r_{min})$	n^{α}_{β}	n^{β}_{α}
1:1:29 Closed A					
$H_C - O_W$	2.295	2.3O	0.175572	1.061890196	0.10985071
$O_C - O_W$	3.405	3.4	0.508798	5.377788235	0.370881947
$O_{CE} - O_W$	3.315	3.4	0.362649	5.362858824	0.369852333
$O - O_W$	3.855	3.4	0.541914	3.665145588	0.252768661
1:1:29 Open A					
$H_C - O_W$	2.295	2.3	0.187153	1.031818627	0.106739858
$O_C - O_W$	3.435	3.4	0.493206	5.217667647	0.359839148
$O_{CE} - O_W$	3.285	3.4	0.389006	5.642661765	0.389149087
$O - O_W$	3.345	3.4	0.519989	3.519942647	0.242754665
1:1:29 Open A					
$H_C - O_W$	2.325	2.3	0.236943	1.33267451	0.13786288
$O_C - O_W$	3.345	3.4	0.4945	5.367129412	0.370146856
$O_{CE} - O_W$	3.255	3.4	0.418119	6.141501471	0.423551826
$O - O_W$	3.885	3.4	0.60976	4.030929412	0.277995132
1:1:61 Closed A					
$H_C - O_W$	2.295	2.3	0.180108	1.289825926	0.063434062
$O_C - O_W$	3.375	3.4	0.496586	6.119441667	0.200637432
$O_{CE} - O_W$	3.315	3.4	0.304314	5.711763889	0.187270947
$O - O_W$	3.285	3.4	0.531974	4.060780556	0.133140346
1:1:61 Open A					
$H_C - O_W$	2.265	2.3	0.172373	1.174214815	0.05774827
$O_C - O_W$	3.345	3.4	0.463179	5.933811111	0.194551184
$O_{CE} - O_W$	3.315	3.4	0.308934	5.539241667	0.181614481
$O - O_W$	3.345	3.4	0.526984	3.993075	0.130920492
1:1:61 Open D					
$H_C - O_W$	2.325	2.3	0.194813	1.315038889	0.064674044
$O_C - O_W$	3.465	3.4	0.473206	5.783755556	0.18963133
$O_{CE} - O_W$	3.255	3.4	0.331723	6.2194	0.203914754
$O - O_W$	3.225	3.4	0.50653	3.899741667	0.127860383

Table C.2: Co-ordination numbers, n^{α}_{β}, for the GPE-water hydrogen bond sites, indicating the average number of β atoms around an α atom. For comparison purposes we used the average of all the O_W-GPE-oxygen first minimum distances, across all six systems to provide the cut-off distance of 3.4 Å for the $O_{CE}/O_C/O-O_W$ co-ordination numbers. The $H_C - O_W$ cut-off distance used, of 2.3 Å is averaged across all six systems.



(a) 1:29 A Closed

(b) 1:29 A Open

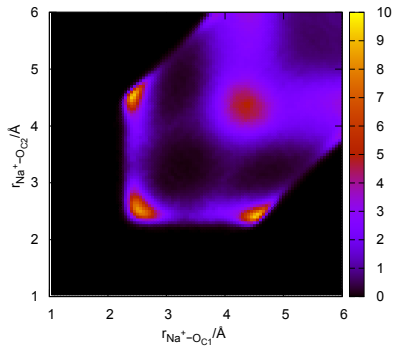
(c) 1:29 D Open

(d) 1:61 A Closed

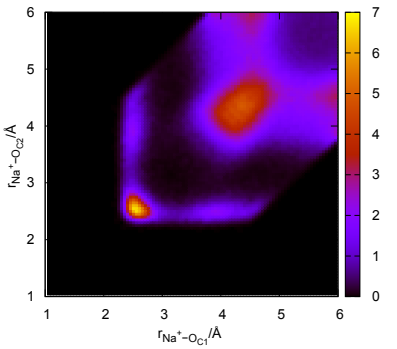
(e) 1:61 A Open

(f) 1:61 D Open

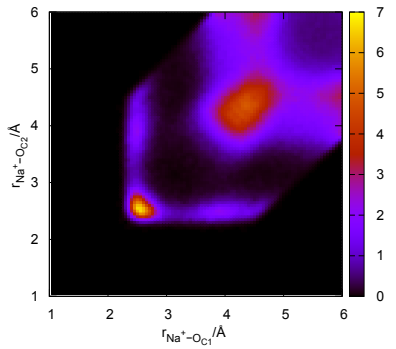
Figure C.16: 2D radial-radial distribution functions for the sodium-GPE interactions $Na^+ - O_{CE}$ in 1:1:29 and 1:1:61 solutions of GPE.



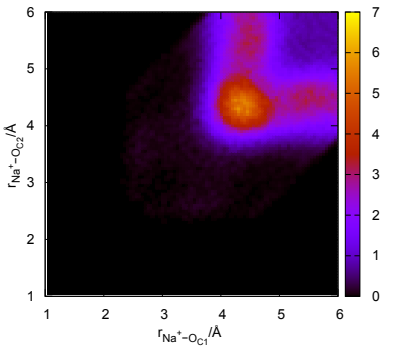
(a) 1:29 A Closed



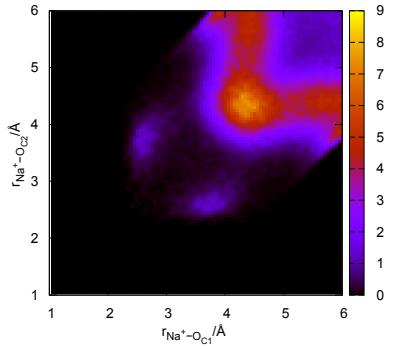
(b) 1:29 A Open



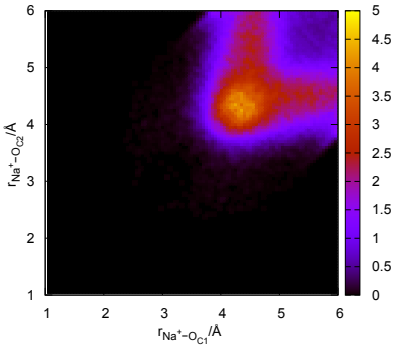
(c) 1:29 D Open



(d) 1:61 A Closed

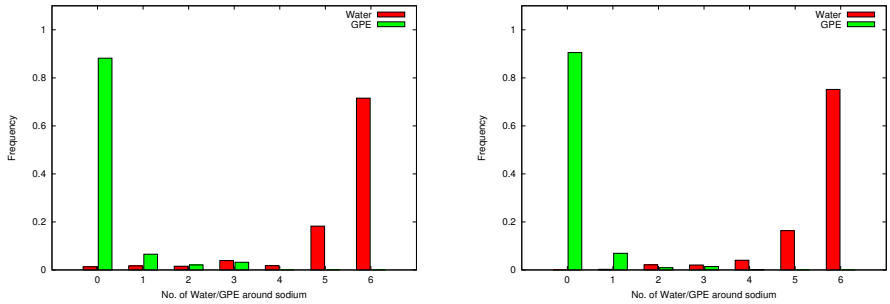


(e) 1:61 A Open



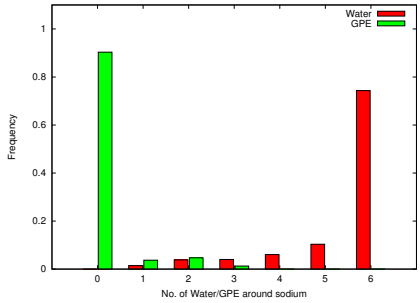
(f) 1:61 D Open

Figure C.17: 2D radial-radial distribution functions for the sodium-GPE interactions $Na^+ - O_{CE}$ in 1:1:29 and 1:1:61 solutions of GPE.

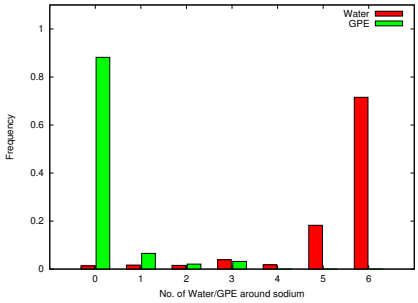


(a) 1:29 A Closed

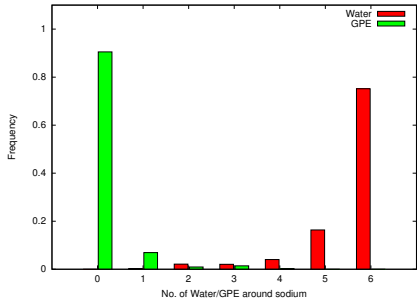
(b) 1:29 A Open



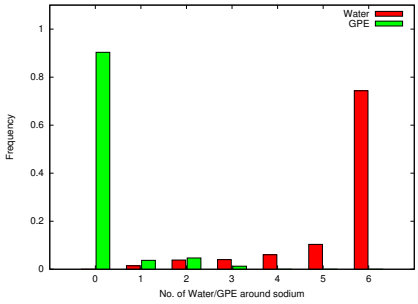
(c) 1:29 D Open



(d) 1:61 A Closed



(e) 1:61 A Open

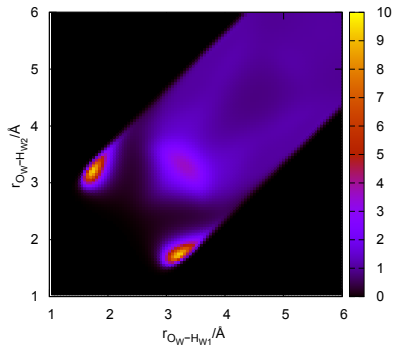


(f) 1:61 D Closed

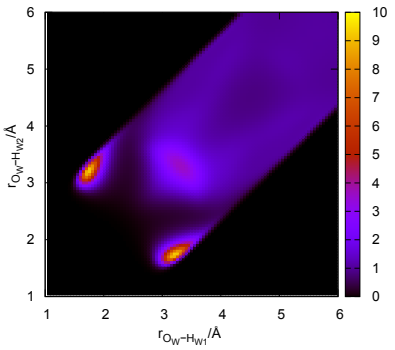
Figure C.18: Histograms showing the proportion of sodium ions with 1,2...6 water(red) and GPE (green) neighbours as a fraction of the total number of sodium ions in the system.

Interaction	r_{min} (Å)	r_{cut} (Å)	$g(r_{min})$	n_β^α	n_α^β
1:1:29 Closed A					
$O_{CE} - Na^+$	3.075	3.1	0.21758	0.084722059	0.169444118
$O_C - Na^+$	3.225	3.1	0.305972	0.169573529	0.339147059
$O - Na^+$	3.285	3.1	0.220313	0.097657353	0.195314706
$O_W - Na^+$	3.075	3.1	0.0834018	0.376061258	10.90577647
$O_W - O_W$	3.825	3.7	1.07141	5.670460649	5.670460649
1:1:29 Open A					
$O_{CE} - Na^+$	3.045	3.1	0.204028	0.105219118	0.210438235
$O_C - Na^+$	3.015	3.1	0.166641	0.069583824	0.139167647
$O - Na^+$	3.135	3.1	0.094702	0.047120588	0.094241176
$O_W - Na^+$	3.045	3.1	0.0893408	0.383228905	11.11363824
$O_W - O_W$	3.825	3.7	1.06573	5.615043509	5.615043509
1:1:29 Open A					
$O_{CE} - Na^+$	3.045	3.1	0.180363	0.144394118	0.288788235
$O_C - Na^+$	3.135	3.1	0.114456	0.073891176	0.147782353
$O - Na^+$	3.225	3.1	0.116152	0.052245588	0.104491176
$O_W - Na^+$	3.105	3.1	0.0741394	0.377694219	10.95313235
$O_W - O_W$	3.735	3.7	1.07653	5.644633469	5.644633469
1:1:61 Closed A					
$O_{CE} - Na^+$	2.995	3.1	0.00628798	0.000352941	0.000705882
$O_C - Na^+$	2.995	3.1	0.0278468	0.001092647	0.002185294
$O - Na^+$	3.165	3.1	0.0258401	0.001875	0.00375
$O_W - Na^+$	3.105	3.1	0.0697819	0.212184584	6.153352941
$O_W - O_W$	3.615	3.7	0.981995	6.564740061	6.564740061
1:1:61 Open A					
$O_{CE} - Na^+$	2.895	3.1	0.00282654	0.000120588	0.000241176
$O_C - Na^+$	3.075	3.1	0.120255	0.006970588	0.013941176
$O - Na^+$	2.865	3.1	0.0182783	0.000735294	0.001470588
$O_W - Na^+$	3.105	3.1	0.0712703	0.212098479	6.150855882
$O_W - O_W$	3.615	3.7	0.971033	6.503189047	6.503189047
1:1:61 Open D					
$O_{CE} - Na^+$	2.955	3.1	0.011567	0.000247059	0.000494118
$O_C - Na^+$	2.865	3.1	0.0056793	0.000332353	0.000664706
$O - Na^+$	3.045	3.1	0.0083795	0.001925	0.00385
$O_W - Na^+$	3.075	3.1	0.0686978	0.212393611	6.159414706
$O_W - O_W$	3.735	3.7	1.0047	6.686606187	6.686606187

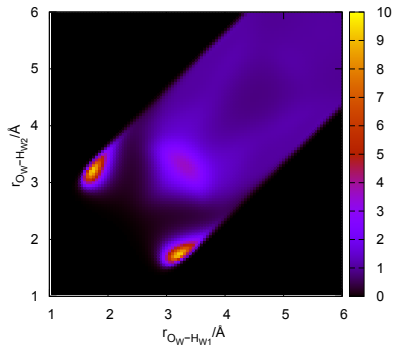
Table C.3: Co-ordination numbers, n_β^α, for sodium co-ordinating with GPE and water, and the water-water ($O_W - O_W$) co-ordinatnion numbers, indicating the average number of β atoms around an α atom. For comparison purposes the cut-off distance used to obtain the $O_{CE}/O_C/O - Na^+$ occupation numbers is taken as the average of r_{min} of all the oxygen-sodium first minimum distances across all six systems, resulting in a cut-off distance of 3.1 Å. The $O_W - O_W$ cut-off distance of 3.7 Å is averaged across all six systems.



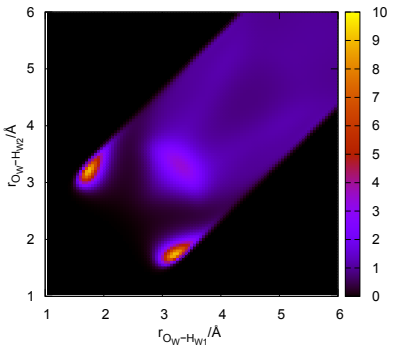
(a) 1:29 A Closed



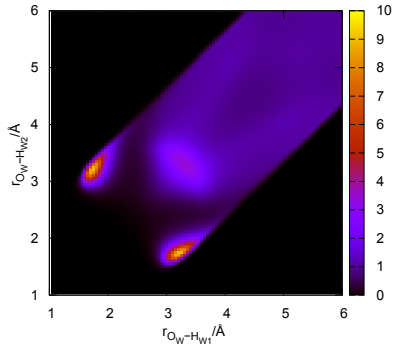
(b) 1:29 A Open



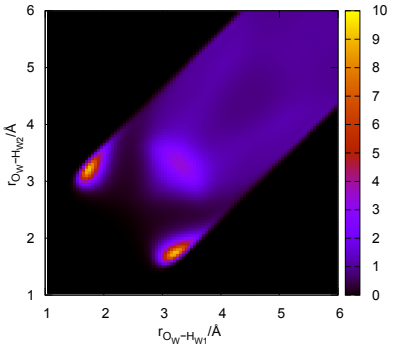
(c) 1:29 D Open



(d) 1:61 A Closed



(e) 1:61 A Open



(f) 1:61 D Open

Figure C.19: 2D radial-radial distribution functions for the water-water interactions $O_W - H_W$ in 1:1:29 and 1:1:61 solutions of GPE.

Appendix D

Sodium Van der Waals Interactions

D.1 Introduction

In recent years there has been a great deal of attention focused on the simulations of aqueous-ion solutions [6, 20, 67, 68, 134, 143, 144, 145, 146, 147, 148, 149, 150]. While it is not the primary aim of the current work to undertake a full survey of potentials, both the glutamate and GPE solutions contain sodium as a counter-ion in the system. Thus in this appendix we consider the effect of the different parameterisations of sodium ions on the system.

In classical simulations, inter-atomic potentials generally include electrostatic interactions and a Lennard-Jones potential. While there are alternative approaches, including the most recent work on polarisable classical potentials [151], empirical potentials with fixed Lennard-Jones parameters currently provide the most efficient and effective method of modelling ions. The charge and mass of ions is easily defined, and thus the electrostatic interaction is easily determined. The remaining parameters come from the Lennard-Jones interaction;

$$U(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \;=\; \left(\frac{A}{r_{ij}^{12}}\right) - \left(\frac{B}{r_{ij}^{6}}\right). \tag{D.1}$$

There are many different potentials for ions in water, all producing slightly different structures, energetic properties and fits to experimental data. These

potentials are developed to optimise the ability of the potential to characterise a particular set of properties, such as the variation in the heat capacity of a system with temperature. However, this often results in other properties being poorly described. Determining the quality of a model is not as simple as comparing with experiment; the quality of the experimental data needs to determined as well as the importance of the statistics used to compare models to experiments. Models are often validated by comparing thermodynamic and structural properties, most commonly the Gibb's free energy of hydration (the difference in energy after the removal of a water molecule from the hydration shell of an ion). Once thermodynamic properties have been fitted, structural features such as radial distribution functions are considered. However, it should be noted that such structural features are not always available experimentally. In addition, to this Jensen & Jorgensen attempted to produce potentials for ions that are consistent across various sizes of ions [152], as well as fitting key thermodynamic properties of experiments.

In the following discussion we consider the energetic and structural properties of some potential models for sodium ions.

D.2 Forcefield parameters

There are many different parametrisations for ions in Molecular Dynamics simulations [6, 68, 71, 153, 132, 143, 152, 144, 154, 155] (although this is not an exhaustive list it is representative of the commonly employed potentials). As with the different empirical models of water, each potential attempts to replicate a different set of experimental results, each in a different context, with different external conditions, and thus there is no 'silver-bullet' for modelling ions in solution. Another problem, encountered in this work, and mentioned by Patra *et al.* [6], is the inconsistency between available sources of the parameters for particular models, where the same model, available from multiple sources, may contain slightly different parameters.

Of particular relevance to our work is the difference between CHARMM parameters. Patra *et al.* noted that in the original CHARMM22 release there is no parameterisation for sodium, unless the X-Plor version 3.851, available at `http://atb.csb.yale.edu/xplor/`, is used, which is labelled CHARMM22 and

includes ion parameters (hereafter referred to as X-Plor/CHARMM22). In setting up the force-fields for the glutamate solutions modelled with DLPOLY_2, and used in obtaining the results discussed in chapter 3, we found that the CHARMM22 files available from the Mackerell website [83] *do* include ion parameters. However, the Lennard-Jones parameters for the sodium ion are the same when downloaded from the Mackerell website for CHARMM22 and CHARMM27 (see files par_all22_prot.inp and par_all27_prot_lipid.prm), but different from the X-Plor/CHARMM22 potentials. Hereafter this set of potentials is referred to as CHARMM22/27

In table D.1 we list the Lennard-Jones parameters, ϵ and σ, from various emprical force-fields for sodium ions, including those used in the work in this thesis. This includes the two different versions of the CHARMM parameters, from X-Plor and from the MacKerell Website. The X-Plor/CHARMM22 parameters for sodium are the same as those used in the PINY_MD classical simulations of glutamate and GPE in this work, and the CHARMM22/27 parameters are the same as those used in the DLPOLY_2 simulations. It should be noted at this point that it is not just the ions in the solution that are affected by the different potentials, but also the ion-solute and ion-solvent interactions.

Forcefield	ϵ[kcal/mol]	σ[Å]
Gromacs [155]	0.0148	2.5752
X-Plor/CHARMM22 [156]	0.1000	2.7297
CHARMM22/27 [154]	0.0469	2.4299
Amber 1999 [132]	0.0028	3.3284
OPLS-AA [153]	0.0028	3.3304
Smith-1994 [68]	0.1001	2.3500
Jensen-2006 [152]	0.0005	4.07

Table D.1: Force-field parameters for sodium, ϵ and σ, for the Lennard-Jones potential used in non-polarisable empirical force-fields.

Patra *et al* performed a comparison of six of the force-fields listed for solutions of Na-Cl [6]. To enable comparison to experiment they considered the Gibbs free energy of hydration;

$$F = -k_B T \; ln \left\langle exp\left(-\frac{U}{k_B T}\right)\right\rangle \tag{D.2}$$

where k_B is the Boltzmann coefficient, T is the temperature and U is the potential energy of the system. Jensen & Jorgensen also calculated the Gibbs free enegy of hydration for their system [152], but this is for a single sodium ion surrounded by water and thus is not directly comparable to the total Gibbs free energy for Na-Cl. Numerically computed Gibb's free energies by Patra *et al.* are shown in table D.2 [6]. All the computed values are more negative than the experimental value obtained. Patra *et al.* notes that this could be because the simulations either over-estimate mutual interaction, or that water shielding, where water molecules reduce the ability for other water molecules to be perturbed by the ion, is under-estimated. They found the closest agreement with experimental values by using the X-Plor/CHARM22 sodium potential and the SPC/E water potential.

Forcefield	SPC/E kJ/mol	SPC kJ/mol	TIP4P kJ/mol	TIP3P kJ/mol
Gromacs	-810	-815	-797	-819
X-Plor/CHARMM22	-741	-747	-747	-750
CHARMM22/27	-829	-835	-832	-837
Amber 1999	-750	-757	-758	-761
OPLS-AA	-778	-786	-784	-790
Smith-1994	-746	-751	-752	-755
Experimental	-722			

Table D.2: Gibbs free energy of hydration for a Na-Cl infinite dilution system for each of the force-fields. Data from ref [6].

Jensen & Jorgensen produce free energy of hydration of 428.86 kJ/mol for just the sodium ion (approximately half the value described in table D.2, as the Cl ion is not included). In comparison, the experimental value for sodium quoted by Patra *et al.* is -375 kJ/mol. Thus as with the results in table D.2 this parametrisation also produces a far more negative Gibbs free energy.

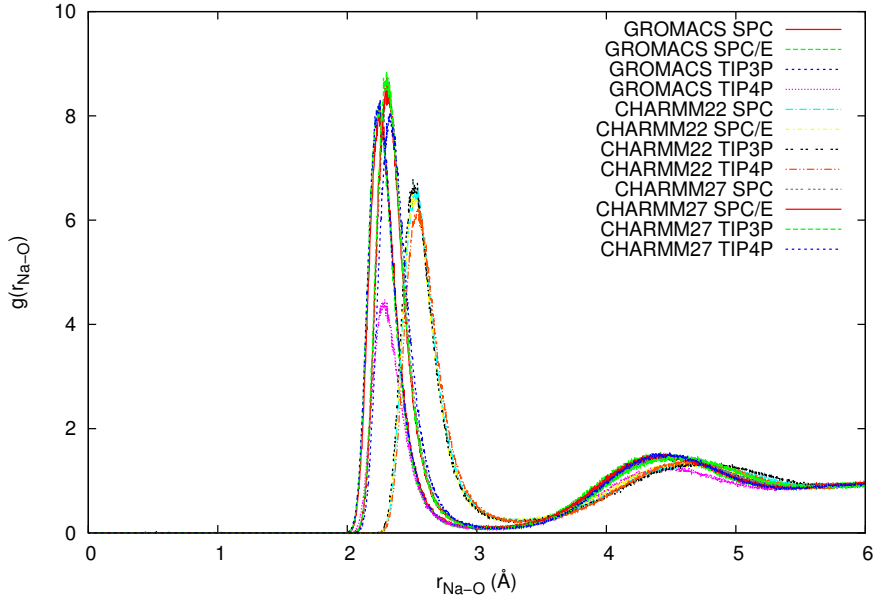
In summary, all the commonly used sodium empirical potentials consistently estimate a more negative Gibbs free energy compared to the experimental values. However, given the nature of simulation it is often of more direct relevance to consider the properties that simulations aim to elucidate, and thus evaluate the potentials accordingly. Radial distributions functions are commonly used to understand the structure of modelled systems. In figure D.1 we show the RDFs for

the sodium-ion to water-oxygen (Na-OW) interaction obtained by Patra *et al.* for Gromacs, X-Plor/CHARMM22 and CHARMM22/27 potentials, figure D.1(a), and for Amber 1999, OPLS-AA and Smith 1994, figure D.1(b). In each instance the data for the RDFs was obtained from the supplementary material of Patra's work for the Nose-Hoover thermostat with the TIP3P, TIP4P, SPC and SPC/E water potentials. For clarity we have spread the data across two plots.

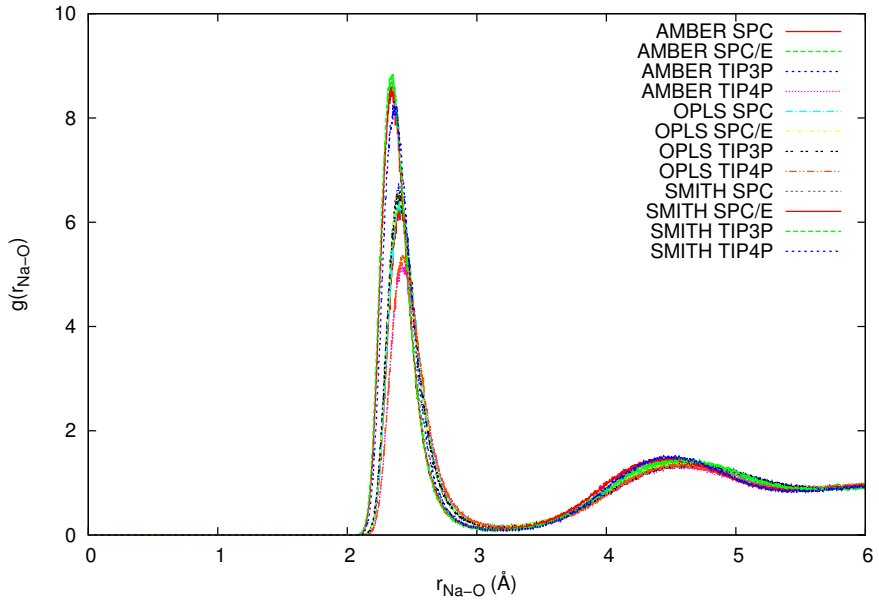
Here we see that there is variation in the position and intensity of the first peak in the RDFs in figure D.1. Patra *et al.* determine that the Gibbs free energies are reproduced 'sufficiently well', but emphasise that, as RDFs cannot be obtained directly from experiment, that there is no correct position and intensity for the peaks in the RDFs, thus suggesting that all of these are equally valid.

To enable further comparison figure D.2 shows the RDFs for all six forcefields with the SPC/E water potential considered by Patra *et al.* with the addition of our classical RDFs using CHARMM22/27 (glu study in DLPOLY_2) and X-Plor/CHARMM22 (as used in PINY_MD simulations of glu for parallel tempering and classical GPE). We see here that the X-Plor/CHARMM22 values from Patra's work and from the Glutamate simulations carried out in PINY_MD (for the Parallel Tempering work of this thesis) lie in an almost exact match, as expected as they used the same sodium potential. The simultations carried out using DLPOLY_2 for the aqueous glutamate solution using CHARMM22/27 (ref [83]) are in closer agreement with Patra's data for CHARMM27, again as expected. Of most interest is the RDF plot for the GPE system NA-OW RDF. This uses X-Plor/CHARMM22, however the position of the peak coincides most closely with the CHARMM22/27 and Amber plots from Patra's work. It should be noted that drawing too many conclusions from comparing the glutamate and GPE solution data with the Na-Cl solutions by Patra *et al.* is misleading as the presence of the large biological molecules perturbs the rest of the solution. Indeed, there could be other valid reasons for the displacement of the peak rather than just the ion parameterisations.

In table D.3 we give the position of the first maximum of the Na-OW (water-oxygen atom) RDF and, where possible, the intensity of that maximum, for the work by Patra *et al.*, a Car-Parrinello (CPMD) study on Na-Cl solutions by Khalack *et al.* [7], and for an Na^+ ion in water by Jensen & Jorgensen [152]. We are unable to provide a $g(r_{\max})$ for Jensen & Jorgensen's work as this is



(a) GROMACS, CHARMM22/X-Plor, CHARMM27



(b) AMBER, OPLS, SMITH

Figure D.1: RDFs obtained by Patra *et al.* for the interaction between sodium and water-oxygen atoms [6]. CHARMM22 refers to the X-Plor/CHARMM22 potential for sodium and CHARMM27 is the CHARMM22/27 potential.

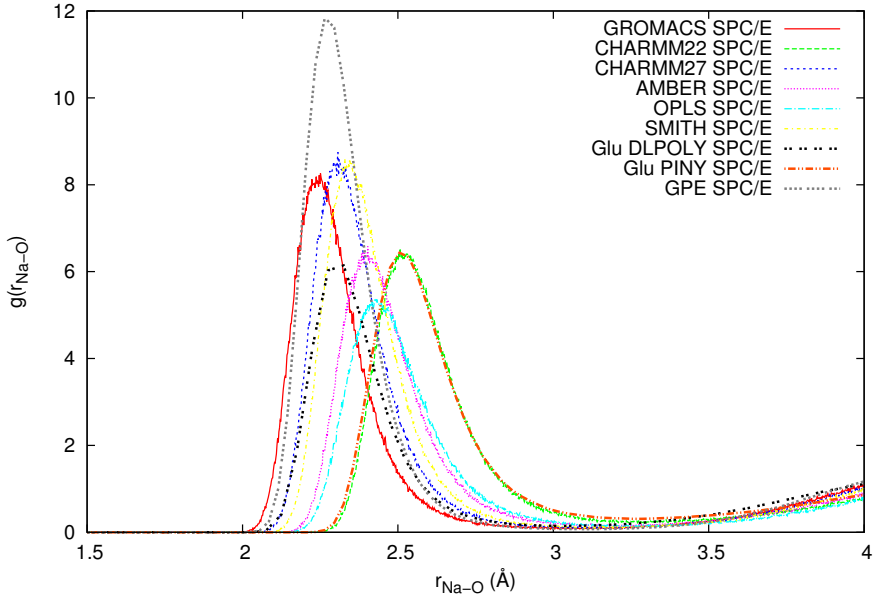


Figure D.2: RDFs obtained by Patra *et al.* [6] for the interaction between sodium and water-oxygen atoms for the X-Plor/CHARMM22 with SPC/E water, and the RDFs obtained in this thesis from the work on glutamate using CHARMM22/27 (DLPOLY_2) and the work on Glu and GPE using X-Plor/CHARMM22 (PINY_MD).

not provided in their paper. Again we include the data from our own solutions, although, as mentioned above, comparisons between our work and solutions of Na-Cl should be limited. We have not included our CPMD study in this data as the Na-OW RDFs were not statistically significant enough to give meaningful results due to the limited number of ions in the system and the proximity to the glutamate molecules.

Forcefield	$r_{\max}$ position (Å)	$g(r_{\max})$
Gromacs	2.247	8.22176
X-Plor/CHARMM22	2.507	6.50624
CHARMM22/27	2.307	8.74773
Amber 1999	2.403	6.58016
OPLS-AA	2.391	6.59915
Smith-1994	2.329	8.60216
CPMD (Khalack)	~ 2.1	~ 7.1
Jensen-2006	2.50	
Glu (CHARMM22/27 - DLPOLY)	2.325	6.141
Glu (X-Plor/CHARMM22 - PINY)	2.505	6.43515
GPE (X-Plor/CHARMM22 - PINY)	2.265	11.8441

Table D.3: First peak positions and intensity for empirical potentials with SPC/E water by Patra *et al.* [6], from a Car-Parinello study of Na-Cl in water by Khalack *et al.* [7] and, from the classical glutamate and GPE simulations.

The positions and intensities of the first minima show significant variation, even in work that is closely comparable, such as in the data by Patra *et al.* where the systems were all subjected to the same simulation parameters. However, all of the systems over-estimate the radial position of the first peak, and therefore the position of the first occupation shell when compared to the CPMD study by Khalack *et al.*. As Car-Parrinello MD is a first-principles approach to simulation, in some respects this should provide the most accurate simulation. However, as discussed in the main body of this work, first-principles simulations are still flawed due to the inherent approximations in the description of the functional and the limited system sizes.

The RDFs for the two glutamate simulations, using the CHARMM22/27 (DLPOLY_2) potential and the X-Plor/CHARMM22 (PINY_MD) potential produce peaks of similar intensity, but at different radial positions (figure D.2) However, the GPE RDF, simulated using the X-Plor/CHARMM22 potential

produces a peak with a much greater intensity, but at a smaller radius than glutamate X-Plor/CHARMM22 RDF, and similar to the position of the glutamate CHARMM22/27 RDF. This indicates that for this work the other contents of the system have a greater impact than the actual parameterisation of the sodium ion used. Thus any comparisons considered between solutions should take this into consideration.

D.3 Conclusion

There are many different empirical potentials available for the study of ions in solution. In this thesis we have found that potentials that appear to be the same often differ, depending on their exact source. However, this summary of previous work on sodium ion solutions has shown that the parameters, provided they are valid and replicate experimental results, are not the primary factor in the different structures obtained from simulation results. We note that the large GPE molecule seems to affect the structure of the sodium-water interactions more than the using different potentials for the sodium ion. This should therefore be taken into consideration in any work done on these systems. In particular if cross-system comparison is required, which is often a major resource in the study of biologically-relevant systems, simulations that use exactly the same parameters, as well as simulations using different parameters should be considered before conclusions on the structure are drawn.

Bibliography

[1] Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Oxford University Press, 1987).

[2] Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Research* **28**, 235–242 (2000).

[3] McLain, S. E., Soper, A. K. & Watts, A. Structural studies on the hydration of l-glutamic acid in solution. *J. Phys. Chem. B* **110**, 21251–21258 (2006).

[4] Collis, A. B., Tulip, P. R. & Bates, S. P. Structure and bonding of aqueous glutamic acid from classical molecular dynamics simulations. *Phys Chem Chem Phys* **12**, 5341–5352 (2010).

[5] Wang, M., Shan, L. & Wang, J. PBD ID: 2FXY; Two peptide fragments G55-172 and K97-A109 from Staphylococcal nuclease exhibit different behaviors in conformational preferences for helix formation. *Biopolymers* **83**, 268–279 (2006).

[6] Patra, M. & Karttunen, M. Systematic comparison of force fields for microscopic simulations of NaCl in aqueous solutions: Diffusion, free energy of hydration, and structural properties. *Journal of Computational Chemistry* **25**, 678–689 (2004).

[7] Khalack, J. & Lyubartsev, A. Car-parrinello molecular dynamics simulations of na^{+}-cl^{-} ion pair in liquid water. *Condensed Matter Physics* **7**, 683–698 (2004).

[8] Hill, T. On Steric Effects. *Journal of Chemical Physics* **14**, 465 (1946).

[9] Westheimer, F. & Mayer, J. The theory of the racemization of optically active derivatives of diphenyl. *Journal of Chemical Physics* **14**, 733–738 (1946).

[10] Dill, K. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).

[11] Israelachvili, J. & Wennerström, H. Role of hydration and water structure in biological and colloidal interactions. *Nature* **379**, (1996).

[12] Ball, P. Water as an active constituent in cell biology. *Chemical Reviews* **108**, 74–108 (2008).

[13] Tarek, M. & Tobias, D. J. The dynamics of protein hydration water: A quantitative comparison of molecular dynamics simulations and neutron-scattering experiments. *Biophysical Journal* **79**, 3244 – 3257 (2000).

[14] Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640–647 (2005).

[15] Daidone, I., Ulmschneider, M. B., Di Nola, A., Amadei, A. & Smith, J. C. Dehydration-driven solvent exposure of hydrophobic surfaces as a driving force in peptide folding. *Proc. Natl. Acad. Sci.* **104**, 15230–15235 (2007).

[16] Tulip, P. R. *et al.* Conformational Plasticity in an HIV-1 Antibody Epitope. *Journal of Physical Chemistry B* **114**, 7942–7950 (2010).

[17] Chan, D. Y. C., Mitchell, D. J., Ninham, B. W. & Pailthorpe, B. A. *Chapter 5: Solvent Structure and Hydrophobic Solutions*, vol. 6 (Plenum press, 1979).

[18] Carey, D. M. & Kroenowski, G. M. Measurement of the Raman spectrum of liquid water. *J. Chem. Phys.* **108**, 2669–2675 (1998).

[19] Bukowski, R., Szalewicz, K., Groenenboom, G. C. & van der Avoird, A. Predictions of the properties of water from first principles. *Science* **315**, 1249–1252 (2007).

[20] Dill, K., Truskett, T., Vlachy, V. & Hribar-Lee, B. Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure* **34**, 173–199 (2005).

[21] Luzar, A. Water hydrogen-bond dynamics close to hydrophobic and hydrophilic groups. *Faraday Discussions* **103**, 29–40 (1996). Meeting on Hydration Processes in Biological and Macromolecular Systems, Sheffield, England, Apr 01-03, 1996.

[22] Csajka, F. S. & D., C. Transition pathways in a many-body system: Application to hydrogen-bond breaking in water. *J. Chem. Phys.* **109**, 1125–1133 (1998).

[23] Havlin, R. *et al.* An experimental and theoretical investigation of the chemical shielding tensors of C-13(alpha) of alanine, valine, and leucine residues in solid peptides and in proteins in solution. *Journal of the Americal Chemical Society* **123**, 10362–10369 (2001).

[24] Rossi, S., Lo Nostro, P., Lagi, M., Ninham, B. W. & Baglioni, P. Specific anion effects on the optical rotation of alpha-amino acids. *Journal of Physical Chemistry B* **111**, 10510–10519 (2007).

[25] Sterpone, F., Stirnemann, G., Hynes, J. T. & Laage, D. Water Hydrogen-Bond Dynamics around Amino Acids: The Key Role of Hydrophilic Hydrogen-Bond Acceptor Groups. *Journal of Physical Chemistry B* **114**, 2083–2089 (2010).

[26] McLain, S. E., Soper, A. K. & Watts, A. Water structure around dipeptides in aqueous solutions. *European Biophysics Journal with Biophysics Letters* **37**, 647–655 (2008). Workshop on Proteins at Work 2007, Perugia, Italy, May 28-30, 2007.

[27] Allison, S. K. *Intermolecular structure and dynamics of aqueous* N-*methylacetamide*. Ph.D. thesis, University of Edinburgh (2007).

[28] Stevens, F. J. & M., S. Computer simulation of protein self-association during small-zone gel filtration. estimation of equilibrium constants. *Biochem J.* **195**, 213–219 (1981).

[29] Prabhu, N. & Sharp, K. Protein-solvent interactions. *Chem Rev.* **106**, 1616–1623 (2006).

[30] Thar, J., Reckien, W. & Kirchner, B. Car-Parrinello molecular dynamics simulations and biological systems. In *Atomistic approaches in modern biology: from quantum chemistry to molecular simulations*, vol. 268 of *Topics in current chemistry*, 133–171 (Springer-Verlag Berlin, Heidelberger Platz 3, D-14197 Berlin, Germany, 2007).

[31] Whitfield, T. W., Martyna, G. J., Allison, S. K., Bates, S. P. & Crain, J. Liquid nma: A suprisingly realistic model for hydrogen bonding motifs in proteins. *Chem. Phys. Lett.* **414**, 210–214 (2005).

[32] Allison, S. K., Bates, S. P., Crain, J. & Martyna, G. J. Solution structure of the aqueous model peptide n-methylacetamide. *J. Phys. Chem. B* **110**, 21319–21326 (2006).

[33] Whitfield, T. W. *et al.* Structure and hydrogen bonding in neat n-methylacetamide: Classical molecular dynamics and raman spectroscopy studies of a liquid of peptidic fragments. *J. Phys. Chem. B* **110**, 3624–3637 (2006).

[34] Tulip, P. R. & Bates, S. P. Peptide aggregation and solvent electrostriction in a simple zwitterionic dipeptide via molecular dynamics simulations. *J. Chem. Phys.* **131**, 015103 (2009).

[35] Richardson, J. S. & Richardson, D. C. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci.* **99**, 2754–2759 (2002).

[36] Shigeri, Y., Seal, R. & Shimamoto, K. Molecular pharmacology of glutamate transporters, EAATs and VGLUTs. *Brain Research Reviews* **45**, 250–265 (2004).

[37] Manev, H., Favaron, M., Guidotti, A. & Costa, E. Delayed increase of Ca2+ influx elicited by glutamate: role in neuronal death. *Mol Pharmacol* **36**, 106–112 (1989).

[38] Hynd, M., Scott, H. & Dodd, P. Glutamate-mediated excitotoxicity and neurodegeneration in Alzheimer's disease. *Neurochemistry International* **45**, 583–595 (2004).

[39] Renton, A. If msg is so bad for you, why doesn't everyone in asia have a headache? (2005). URL `http://www.guardian.co.uk/lifeandstyle/2005/jul/10/foodanddrink.features` [Accessed: 17 January 2011].

[40] Olney, J. W. Brain lesions, obesity, and other disturbances in mice treated with monosodium glutamate. *Science* **164**, 719–721 (1969).

[41] Ohguro, H. *et al.* A high dietary intake of sodium glutamate as flavoring (ajinomoto) causes gross changes in retinal morphology and function. *Experimental Eye Research* **75**, 307 – 315 (2002).

[42] Beyreuther, K. *et al.* Consensus meeting: monosodium glutamate - an update. *Eur. J. Clinial Nutrition* **61**, 928 (2007).

[43] Freeman, E. Reconsidering the effects of monosodium glutamate: a literature review. *J Am Acad Nurse Pract* **18**, 482–486 (2006).

[44] Leenders, E. J. M., Bolhuis, P. G. & Meijer, E. J. Microscopic picture of the aqueous solvation of glutamic acid. *J. Chem. Theory Comput.* **4**, 898–907 (2008).

[45] Daub, C. D., Leung, K. & Luzar, A. Structure of Aqueous Solutions of Monosodium Glutamate. *J. Phys. Chem. B* **113**, 7687–7700 (2009).

[46] Degtyarenko, I., Jalkanen, K. J., Gurtovenko, A. A. & Nieminen, R. M. The aqueous and crystalline forms of l-alanine zwitterion. *Journal of Computational and Theoretical Nanoscience* **5**, 277–285 (2008). URL `http://www.ingentaconnect.com/`.

[47] Ding, Y. & Krogh-Jespersen, K. The Glycine zwitterion does not exist in the gas-phase - results from a detailed ab initio electronic-structure study. *Chemical Physics Letters* **199**, 261–266 (1992).

[48] Jensen, J. & Gordon, M. On the number of water-molecules necessary to stabilize the glycine zwitterion. *Journal of the American Chemical Society* **117**, 8159–8170 (1995).

[49] Raman, S. S., Parthasarathi, R., Subramanian, V. & Ramasami, T. Role of aspartic acid in collagen structure and stability: A molecular dynamics investigation. *Journal of Physical Chemistry B.* **110**, 20678–20685 (2006).

[50] Bostrom, M., Williams, D. & Ninham, B. Specific ion effects: The role of co-ions in biology. *Europhysics Letters* **63**, 610–615 (2003).

[51] Hermansson, K. & Wojcik, M. Water exchange around Li+ and Na+ in LiCl(aq) and NaCl(aq) from MD simulations. *J. Phys. Chem. B* **102**, 6089–6097 (1998).

[52] Guan, J., Waldvogel, H., Faull, R., Gluckman, P. & Williams, C. The effects of the n-terminal tripeptide of insulin-like growth factor-1, glycine-proline-glutamate in different regions following hypoxic-ischemic brain injury in adult rats. *Neuroscience* **89**, 649 – 659 (1999).

[53] Guan, J. *et al.* N-terminal tripeptide of IGF-1 (GPE) prevents the loss of TH positive neurons after 6-OHDA induced nigral lesion in rats. *Brain Research* **859**, 286–292 (2000).

[54] Alexi, T. *et al.* The IGF-I amino-terminal tripeptide glycine-proline-glutamate (GPE) is neuroprotective to striatum in the quinolinic acid lesion animal

model of Huntington's disease. *Experimental Neurology* **159**, 84–97 (1999). 2nd Conference of Cellular and Molecular Treatments of Neurological Diseases (CMT), Cambridge, Massachusetts, OCT, 1998.

[55] R., S. V. *et al.* Characterization of somatomedins from human-fetal brain - identification of a variant form of insulin-like growth factor 1. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 4904–4907 (1986).

[56] C., C. *et al.* Isolation and charaterization of variant IGF-1 as well as IGF-2 from adult human-brain. *FEBS Letters* **201**, 46–50 (1986).

[57] Sandberg-Nordqvist, A. C., Stahlbom, P. A., Lake, M. & Sara, V. R. Characterization of two cdnas encoding insulin-like growth factor 1 (igf-1) in the human fetal brain. *Brain Res Mol Brain Res* **12**, 275–7 (1992).

[58] Sara, V. R. *et al.* The Biological Role of Truncated Insulin-like Growth Factor-1 and the Tripeptide GPE in the Central Nervous System. In Raizada, MK and Leroith, D (ed.) *The Role of Insulin-Like Growth Factors in the Nervous Systems*, vol. 692 of *Annals of the New York Academy of Sciences*, 183–191. New York Acad Sci (New York Acad Sciences, 1993).

[59] Wang, J., Cai, K. & Ma, X. Ultrafast Structural Dynamics of Biomolecules Examined by Multiple-Mode 2D IR Spectroscopy: Anharmonically Coupled Motions are in Harmony. *ChemPhysChem* **10**, 2242–2250 (2009).

[60] Richards, W. G. *Chapter 3: Ab Initio Methods and the Study of Molecular Hydration*, vol. 6 (Plenum Pub Corp, 1979).

[61] Rapaport, D. C. Hydrogen bonds in water: Network and organisation lifetimes. *Mol. Phys.* **50**, 1151–1162 (1983).

[62] Belch, A. C. & Rice, S. A. The distribution of rings of hydrogen-bonded molecules in a model of liquid water. *J. Chem. Phys.* **86**, 5676–5682 (1987).

[63] Ohmine, I. & Tanaka, H. Fluctuation, relaxations, and hydration in liquid water. hydrogen-bond rearrangement dynamics. *Chem. Revs.* **93**, 2545–2566 (1993).

[64] Rick, S. W., Stuart, S. J. & Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **101**, 6141–6156 (1994).

[65] Keutsch, F. N. & Saykally, R. J. Water clusters: Untangling the mysteries of the liquid, one moelcule at a time. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10533–10540 (2001).

[66] Megyes, T. *et al.* The structure of aqueous sodium hydroxide solutions: A combined solution x-ray diffraction and simulation study. *J. Chem. Phys.* **128** (2008).

[67] Dang, L., Rice, J. & Kollman, P. The effect of water models on the interaction of the sodium-chloride ion-pair in water - Molecular Dynamics simulations. *Journal of Chemical Physics* **93**, 7528–7529 (1990).

[68] Smith, D. & Dang, L. Computer-Simulations of NACL association in polarizable water. *Journal of Chemical Physics* **100**, 3757–3766 (1994).

[69] Park, S. & Fayer, M. D. Hydrogen bond dynamics in aqueous NaBr solutions. *Proc. Natl. Acad. Sci.* **104**, 16731–16738 (2007).

[70] Mancinelli, R., Botti, A., Bruni, F., Ricci, M. A. & Soper, A. K. Hydration of sodium, potassium, and chloride ions in solution and the concept of structure maker/breaker. *J. Chem. Phys. B* **111**, 13570–13577 (2007).

[71] MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

[72] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926–935 (1983).

[73] Mark, P. & Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).

[74] Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press Inc. U.S., 2001), second edn.

[75] Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide* (Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002).

[76] P., H. J. & R., M. I. *Theory of Simple Liquids; 3rd edition* (Academic Press, Elsevier, London, UK, 2006).

[77] Hoover, W. Canonical dynamics - Equilibrium phase-space distributions. *Physical Review A* **31**, 1695–1697 (1985).

[78] Berendsen, H., Postma, J., Vangunsteren, W. F., Dinola, A. & Haak, J. R. Molecular-dynamics with coupling to an external bath. *Journal of Chemical Physics* **81**, 3684–3690 (1984).

[79] Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé-hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).

[80] Verlet, L. Computer experiments on classical fluids .I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review* **159**, 98–103 (1967).

[81] Levitt, M., Hirshberg, M., Sharon, R., Laidig, K. E. & Daggett, V. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem. B* **101**, 5051–5061 (1997).

[82] Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Chem. Phys.* **91**, 6269–6271 (1987).

[83] MacKerell, A. CHARMM FF parameters (2010). URL `http://mackerell.umaryland.edu/CHARMM_ff_params.html` [Accessed: 6 December 2010].

[84] Nutt, D. R. & Smith, J. C. Molecular dynamics simulations of proteins: Can the explicit water model be varied? *Journal of Chemical Theory and Computation* **3**, 1550–1560 (2007).

[85] Abascal, J. L. F. & Vega, C. A general purpose model for the condensed phases of water: Tip4p/2005. *J. Chem. Phys.* **123**, 234505 (2005).

[86] Mahoney, M. & Jorgensen, W. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *Journal of Chemical Physics* **112**, 8910–8922 (2000).

[87] Guillot, B. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **101**, 219–260 (2002).

[88] Car, R. & Parrinello, M. Unified Approach for Molecular-Dynamics and Density-Functional theory. *Physical Review Letters* **55**, 2471–2474 (1985).

[89] Tuckerman, M. & Parrinello, M. Integrating the Car-Parrinello equations .1. Basic Integration Techniques. *Journal of Chemical Physics* **101**, 1302–1315 (1994).

[90] Tuckerman, M. & Parrinello, M. Integrating the Car-Parrinello equations .2. Multiple time-scale techniques. *Journal of Chemical Physics* **101**, 1316–1329 (1994).

[91] Hutter, J., Tuckerman, M. & Parrinello, M. Integrating the Car-Parrinello equations .3. Techniques for ultrasoft pseudopotentials. *Journal of Chemical Physics* **102**, 859–871 (1995).

[92] Marx, D. & Hutter, J. Ab Initio *Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, 2009).

[93] Payne, M., Teter, M., Allan, D., Arias, T. & Joannopoulos, J. Iterative minimization tehcniques for ab initio total-energy calculations - Molecular-Dynamics and conjugate gradients. *Reviews of Modern Physics* **64**, 1045–1097 (1992).

[94] Becke, A. Density-functional exchange-energy approximation with correct asymptotic-behaviour. *Physical Review A* **38**, 3098–3100 (1988).

[95] Sprik, M., Hutter, J. & Parrinello, M. Ab initio molecular dynamics simulation of liquid water: Comparison three gradient-corrected density functionals. *Journal of Chemical Physics* **105**, 1142–1152 (1996).

[96] Lee, C., Yang, W. & Parr, R. Developlemt of the Colle-Salvetti Correltaion-Energy Formula into a Functional of the Electron-Density. *Physical Review B* **37**, 785–789 (1988).

[97] Geyer, C. J. *Markov Chain Monte Carlo Maximum Likelihood*, 156–163 (American Statistical Association, New York, 1991).

[98] Hansmann, U. H. E. Parallel Tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140–150 (1997).

[99] Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).

[100] Earl, D. J. & Deem, M. W. Parallel Tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **7**, 3910–3916 (2005).

[101] Lei, H. & Duan, Y. Improved sampling methods for molecular simulation. *Current Opinion in Structural Biology* **17**, 187–191 (2007).

[102] Nymeyer, H. & Garcia, A. E. Simulation of the folding equilibrium of α-helical peptide: A comparison of the generalized born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13934–13939 (2006).

[103] Rathore, N., Chopra, M. & de Pablo, J. J. Optimal allocation of replicas in Parallel Tempering simulations. *J. Chem. Phys.* **122**, 024111 (2005).

[104] Trebst, S., Huse, D. A. & Troyer, M. Optimizing the ensemble for equilibration in broad-histogram Monte Carlo simulations. *Phys. Rev. E* **70**, 046701 (2004).

[105] Trebst, S., Troyer, M. & Hansmann, U. H. E. Optimized Parallel Tempering simulations of proteins. *J. Chem. Phys.* **124**, 174903 (2006).

[106] Troitzsch, R. Z., Martyna, G. J., McLain, S. E., Soper, A. K. & Crain, J. Structure of aqueous proline via Parallel Tempering Molecular Dynamics and neutron diffraction. *J. Phys. Chem. B* **111**, 8210–8222 (2007).

[107] Helmut G Katzgraber, D. A. H., Simon Trebst & Troyer, M. Feedback-optimized Parallel Tempering Monte Carlo. *Journal of Statistical Mechanics: Theory and Experiment* **2006**, P03018 (2006).

[108] Nadler, W. & Hansmann, U. H. E. Optimizing replica exchange moves for molecular dynamics. *Phys Rev E* **76**, 057102 (2007).

[109] Martyna, G. J. & Tuckerman, M. E. The PINY_MD(c) simulation package (2008). URL `http://homepages.nyu.edu/~mt33/PINY_MD/PINY.html` [Accessed: 7 February 2008].

[110] Tuckerman, M. E., Yarne, D. A., Samuelson, S. O., Hughes, A. L. & Martyna, G. J. Exploiting multiple levels of parallelism in Molecular Dynamics based calculations via modern techniques and software paradigms on distributed memory computers. *Computer Physics Communications* **128**, 333–376 (2000).

[111] McLain, S. E., Soper, A. K., Terry, A. E. & Watts, A. Structure and hydration of L-proline in aqueous solutions. *J. Phys. Chem. B* **111**, 4568–4580 (2007).

[112] Soper, A. Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water. *Molecular Physics* **99**, 1503–1516 (2001).

[113] Soper, A. Partial structure factors from disordered materials diffraction data: An approach using empirical potential structure refinement. *Physical Review B* **72** (2005).

[114] Troitzsch, R. Z., Tulip, P. R., Crain, J. & Martyna, G. J. A Simplified Model of Local Structure in Aqueous Proline Amino Acid Revealed by First-Principles Molecular Dynamics Simulations. *Biophysical Journal* **95**, 5014–5020 (2008).

[115] Brown, I. D. On the geometry of O-H...O hydrogen bonds. *Acta. Cryst.* **A32**, 24–31 (1976).

[116] Pedersen, B. Geometry of hydrogen-bonds from donor water molecules. *ACTA Crystallographica section B - Structural Science* **B 30**, 289–291 (1974).

[117] Marti, J. Dynamic properties of hydrogen-bonded networks in supercritical water. *Phys. Rev. E* **61**, 449–456 (2000).

[118] Matsumoto, M. & Gubbins, K. Hydrogen bonding in liquid methanol. *Journal of Chemical Physics* **93**, 1981–1994 (1990).

[119] Padro, J., Saiz, L. & Guardia, E. Hydrogen bonding in liquid alcohols: a computer simulation study. *Journal of Molecular Structure* **416**, 243–248 (1997).

[120] Haughney, M., Ferrario, M. & McDonald, I. Molecular-dynamics simulation of liquid methanol. *Journal of Physical Chemistry* **91**, 4934–4940 (1987).

[121] Saiz, L., Padro, J. A. & Guardia, E. Structure and dynamics of liquid ethanol. *J. Phys. Chem. B* **101**, 78–86 (1997).

[122] Padro, J. A., Saiz, L. & Guardia, E. Hydrogen bonding in liquid alcohols: a computer simulation study. *J. Mol. Strct.* **416**, 243–248 (1997).

[123] Guàrdia, E., Martì, J., Padró, J. A., Saiz, L. & Komolkin, A. V. Dynamics in hydrogen bonded liquids: water and alcohols. *J. Mol. Liq.* **96-97**, 3–17 (2002).

[124] Ramachandran, G., Ramakrishnan, C. & Sasisekharan, V. Sterorechemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**, 95–& (1963).

[125] Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–& (1996).

[126] Starr, F. W., Nielsen, J. K. & Stanley, H. E. Hydrogen-bond dynamics for the extended simple point-charge model of water. *Phys. Rev. E* **62**, 579–587 (2000).

[127] Luzar, A. Resolving the hydrogen bond dynamics conundrum. *J. Chem. Phys.* **113**, 10663–10675 (2000).

[128] Nieto-Draghi, C., Hargreaves, R. & Bates, S. P. Structure and dynamics of water in aqueous methanol. *J. Phys. Cond. Mat.* **17**, s3265–s3272 (2005).

[129] Starr, F. W., Nielsen, J. K. & Stanley, H. E. Fast and slow dynamics of hydrogen bonds in liquid water. *Phys. Rev. Lett.* **82**, 2294–2297 (1999).

[130] Andrew Turner. *HPCx Userguide* (2009). URL `"http://www.hpcx.ac.uk/support/documentation/UserGuide/HPCxuser/HPCxuser.html"` [Accessed: 21 July 2009].

[131] Kuharski, R. A. & Rossky, P. J. A quantum mechanical study of structure in liquid H_2O and D_2O. *J. Chem. Phys.* **82**, 5164–5177 (1985). URL `http://link.aip.org/link/?JCP/82/5164/1`.

[132] Cornell, W. D. *et al.* A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179–5197 (1995).

[133] Takahashi, M. *et al.* The role of glutamate transporters in glutamate homeostasis in the brain. *Journal of Experimental Biology* **200**, 401–409 (1997).

[134] Schwegler, E. First-principles simulations of aqueous solutions. In *SciDac 2007: Scientific Discovery Through Advanced Computing*, vol. 78 of *Journal of Physics Conference Series*, U497–U504. US DOE Off Sci; Natl Nucl Security Adm; US NSF (IOP Publishing Ltd., Dirac House, Temple Back, Bristol BS1 6BE, England, 2007). 3rd Annual Scientific Discovery through Advanced Computing Conference (SciDAC 2007), Boston, MA, JUN 24-28, 2007.

[135] Ferrario, M., Haughney, M., McDonald, I. & Klein, M. Molecular-dynamics simulation of aqueous mixtures - methanol, acetone, and ammonia. *Journal of Chemical Physics* **93**, 5156–5166 (1990).

[136] Teixeira, J., Bellissent-Funel, M. C., Chen, S. H. & Dianoux, A. J. Experimental determination of the nature of diffusive motions of water molecules at low temperatures. *Phys. Rev. A* **31**, 1913–1917 (1985).

[137] Nieto-Draghi, C., Avalos, J. & Rousseau, B. Transport properties of dimethyl sulfoxide aqueous solutions. *Journal of Chemical Physics* **119**, 4782–4789 (2003).

[138] Nieto-Draghi, C., Avalos, J., Contreras, O., Ungerer, P. & Ridard, J. Dynamical and structural properties of benzene in supercritical water. *Journal of Chemical Physics* **121**, 10566–10576 (2004).

[139] Kone, A. & Kofke, D. A. Selection of temperature intervals for Parallel-Tempering simulations. *J. Chem. Phys.* **12**, 206101 (2005).

[140] Mohanty, S. & Hansmann, U. H. E. Folding of proteins with diverse folds. *Biophysical Journal* **91**, 3573–3578 (2006).

[141] Nadler, W. & Hansmann, U. H. E. Generalized ensemble and tempering simulations: A unified view. *Phys. Rev. E* **75**, 026109 (2007).

[142] Predescu, C., Predescu, M. & Ciobanu, C. The incomplete beta function law for parallel tempering sampling of classical canonical systems. *Journal of Chemical Physics* **120**, 4119–4128 (2004).

[143] Fennell, C. J., Bizjak, A., Vlachy, V. & Dill, K. A. Ion Pairing in Molecular Simulations of Aqueous Alkali Halide Solutions. *Journal of Physical Chemistry B* **113**, 6782–6791 (2009).

[144] Marcus, Y. Ionic-radii in aqueous solutions. *Chemical Reviews* **88**, 1475–1498 (1988).

[145] Zhang, Z. & Duan, Z. Lithium chloride ionic association in dilute aqueous solution: a constrained molecular dynamics study. *Chemical Physics* **297**, 221–233 (2004).

[146] Hribar, B., Southall, N., Vlachy, V. & Dill, K. How ions affect the structure of water. *Journal of the American Chemical Society* **124**, 12302–12311 (2002).

[147] Degreve, L. & Mazze, F. Molecular simulation of LiCl aqueous solutions. *Molecular Physics* **101**, 1443–1453 (2003).

[148] Degreve, L. & da Silva, F. Detailed microscopic study of 1 M aqueous NaCl solution by computer simulations. *Journal of Molecular Liquids* **87**, 217–232 (2000). 5th Liblice Conference on the Statistical Mechanics of Liquids, Zelezna, Ruda, Czech Republic, Jun 07-12, 1998.

[149] Lyubartsev, A. & Laaksonen, A. Concentration effects in aqueous NaCl solutions. A molecular dynamics simulation. *Journal of Physical Chemistry* **100**, 16410–16418 (1996).

[150] Lyubartsev, A. & Laaksonen, A. Calculation of effective interaction potentials from radial-distribution functions - a reverse Monte-Carlo approach. *Physical Review E* **52**, 3730–3737 (1995).

[151] Lamoureux, G., Roux, B. & MacKerell, A. D. J. A simple polarizable model of water based on classical drude oscillators. *J. Chem. Phys.* **119**, 5185–5197 (2003).

[152] Jensen, K. P. & Jorgensen, W. L. Halide, ammonium, and alkali metal ion parameters for modeling aqueous solutions. *Journal of Chemical Theory and Computation* **2**, 1499–1509 (2006).

[153] Jorgensen, W., Maxwell, D. & TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Amer. Chem. Soc.* **118**, 11225–11236 (1996).

[154] MacKerell, A., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**, 257–265 (2000).

[155] Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **7**, 306–317 (2001).

[156] Brunger, A. T. X-PLOR, Version 3.1. a system for X-Ray crystallography and NMR (1992).

Publications

Collis, A. B. K, Tulip, P. R. & Bates, S. P. Structure and bonding of aqueous glutamic acid from classical Molecular Dynamics simulations. In *Physical Chemistry Chemical Physics*, 2010, **12**, 5341-5352.

Collis, A. B. K. & Bates, S. P. Further investigations into the structure of glutamate from parallel tempering simulations. In preparation.